

Estudi Estadístic dels Ciclons Tropicals



Cristina Lidón Moyano

Agraïments: Agreixo profundament al meu tutor, Josep Lluís Solé,
la seva dedicació i paciència durant tot el procés que ha estat
necessari per a dur a terme aquest treball.

Índex

1. Introducció	Pàg. 1
2. Introducció de les dades	Pàg. 2
2. 1. Conceptes rellevants sobre ciclons	Pàg. 3
2. 2. Manipulació de la base de dades	Pàg. 4
2. 3. Estimació de dades faltants	Pàg. 6
3. Anàlisi de sèries temporals	Pàg. 8
3. 1. Estudi de les sèries anuals	Pàg. 8
3. 2. Estudi de les sèries mensuals	Pàg. 10
4. Índexs PDI, IEI i OEI	Pàg. 15
4. 1. Estudi de la relació entre índexs	Pàg. 16
4. 2. Estimació de la funció de densitat	Pàg. 19
4. 3. Estimació de l' exponent de potència	Pàg. 22
4. 4. Estimació de la funció de densitat amb el mètode Kernel.....	Pàg. 25
4. 5. Test de bondat d' ajust	Pàg. 28
5. Conclusions	Pàg. 32
ANNEXOS	Pàg. 33
BIBLIOGRAFIA	Pàg. 59

1. Introducció

L' Estadística és una ciència que principalment vol extreure informació fiable de les dades i, entre altres coses, modelar els fenòmens quotidians que ens envolten. Amb aquesta idea al cap, la motivació d' aquest treball es estudiar el comportament dels ciclons tropicals. Un primer treball orientat en aquesta direcció és "*Disipación y Energía de los Ciclones Tropicales: Ajustes y Test de Bondad de Ajuste*" de Oliver Planes, escrit al 2012. En el nostre cas, però, aprofundirem més l' anàlisi estadístic de les dades.

D' igual manera que en aquest primer treball citat, utilitzarem unes bases de dades que contenen informació sobre ciclons del Atlàntic i del Pacífic mesurada des de l' inici de cada un dels ciclons registrats fins a la seva extinció, en intervals de sis hores i entendrem cada cicló com un experiment aleatori. La primera part del treball consistirà, doncs, en la manipulació i depuració d' aquestes dues bases de dades, tot aquest procés es pot trobar detallat al segon apartat d' aquest treball.

Per altra banda, l' anàlisi de dades observades en diferents moments del temps, és a dir, de sèries temporals, comporta nous problemes pel que fa tant a la modelització estadística com a l' inferència. La obvia correlació introduïda pel mostreig de punts consecutius en el temps pot restringir severament l' aplicabilitat de molts mètodes estadístics convencionals que sovint depenen de l' hipòtesi d' observacions independents i idènticament distribuïdes.

L' objectiu principal d' aquest treball, doncs, es divideix en dues parts. En primer lloc, es realitzarà l' anàlisi de les sèries del número de ciclons registrats anualment i mensualment tant a l' Atlàntic com al Pacífic, tal com podem veure al tercer apartat d' aquest treball. Intentarem, d' aquesta manera, trobar patrons o estructures en aquestes sèries a tal de poder entendre aquest fenomen atmosfèric, i concretament, la seva aparició. Per altra banda, en meteorologia es coneixen tres índexs que defineixen la força o intensitat dels ciclons, aquests s' anomenen PDI, IEI i OEI. El quart apartat d' aquest treball consisteix en l' estudi a fons d' aquests, estudiant primerament la correlació entre ells, seguit de l' estimació de la seva funció de densitat, aplicant un mètode que podríem anomenar naïve i aplicant el mètode Kernel, el que ens permetrà estimar els paràmetres que defineixen la llei de potencies, i per últim, aplicarem un test de bondat d' ajust per a provar si aquests índexs podrien seguir una distribució de Pareto.

Pel últim cal aclarir que el software utilitzat per a la manipulació de les dades abans esmentades ha estat R a la versió 2.12.2. A l' annex, apartat A7, podem trobar tota la sintaxis utilitzada per a dur a terme aquest treball.

2. Introducció de les dades

Per a la realització d'aquest treball s'utilitzen dues bases de dades que ens proporcionen el "*Regional and Mesoscale Meteorology Branch (RAMMB)*", o més concretament el "*National Hurricane Center (NHC)*", i que podem trobar al següent enllaç,

http://rammb.cira.colostate.edu/research/tropical_cyclones/tc_extended_best_track_dataset/

En aquestes bases de dades trobem informació sobre tempestes tropicals localitzades a l'oceà Atlàntic i al nord Pacífic. Concretament tenim les següents variables en el mateix ordre en el que les introduïm:

- | | |
|---|---|
| <ul style="list-style-type: none">• Número d'identificació• Nom de la tempesta• Mes, dia, hora, any (la informació d'aquestes quatre últimes variables es troba en una de sola)• Latitud (deg N)• Longitud (deg W)• Velocitat màxima del vent (kt)• Pressió central mínima (hPa)• El radi de la velocitat màxima del vent (nm)• El diàmetre de l'ull (nm)• Pressió de l'isòbara més llunyana (hPa)• Radi de l'isòbara més llunyana (nm)• Radi des del centre de la tempesta fins a trobar vents de 34 kt de velocitat en direcció nord-est, sud-est, sud-oest, nord-oest (nm)• Radi des del centre de la tempesta fins a trobar vents de 50 kt de | <ul style="list-style-type: none">velocitat en direcció nord-est, sud-est, sud-oest, nord-oest (nm)• Radi des del centre de la tempesta fins a trobar vents de 64 kt de velocitat en direcció nord-est, sud-est, sud-oest, nord-oest (nm)• Codi del tipus de tempesta, on s'utilitza aquest símbol * per a un sistema tropical (depressió tropical, tempesta tropical o huracà), W per onades tropicals, D per pertorbacions tropicals, S per tempestes subtropicals, E per tempestes extratropicals i L per a residual de baixes pressions• Distància a la massa de terra més propera (km), on la illa Trinitat és la menor àrea considerada terra, en aquesta mesura, valors negatius indiquen que la tempesta es troba sobre terra. |
|---|---|

Per a les bases de dades en estudi, cada una de les línies fa referència a informació sobre les variables anteriors d'una tempesta concreta reportada en una determinada hora. És important remarcar que la base de dades de ciclons de l'oceà Atlàntic recull informació entre els anys 1988 i 2011, però la del nord Pacífic recull informació tan sols entre 2001 i 2011. En total, comptem amb 14384 observacions, on 10222 fan referència a l'Atlàntic, amb 147 ciclons diferents mesurats, i 4162 fan referència al Pacífic, amb 96 ciclons diferents. Tenim, per tant, registrats en total, 243 ciclons diferents.

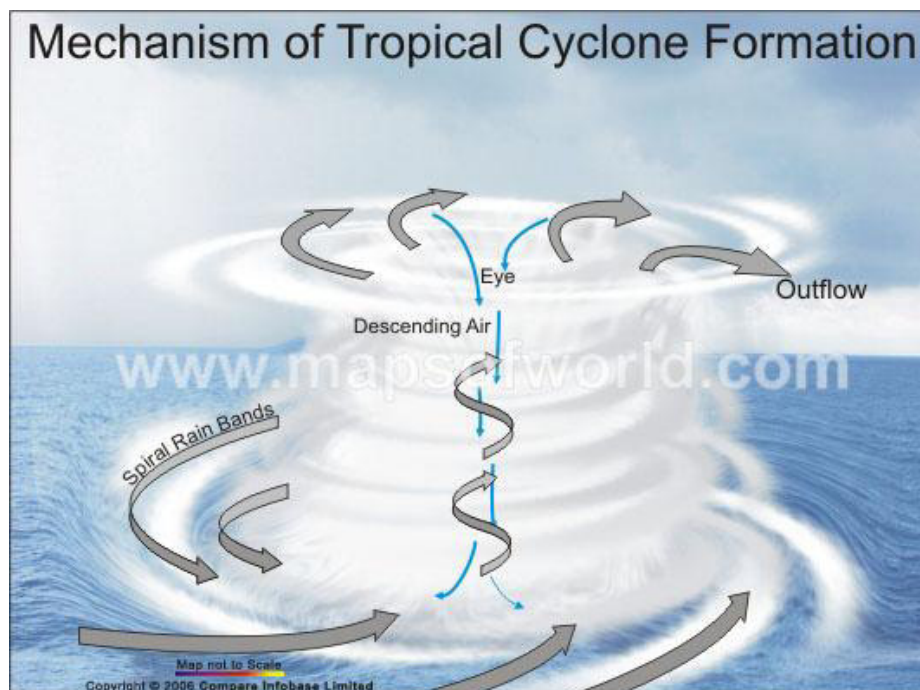
Per últim indicar que el valor numèric -99 que apareix a la base de dades en qualsevol de les variables indica que l' estimació del valor d' aquella variable no va ser possible i, s' entén, per tant, com una dada faltant.

2.1 Conceptes rellevants sobre ciclons

Ja que les nostres dades tracten de tempestes tropicals, pot ser útil comentar breument algunes idees interessants sobre aquest tema.

Comencem aclarint que, quan parlem de **tempestes tropicals**, ens referim a qualsevol pertorbació de l' atmosfera, especialment afectant a la seva superfície, que tenen com a resultat, generalment, forts vents i precipitacions abundants. El terme tropical, fa referència tant a la regió geogràfica on s' originen aquestes tempestes, que acostuma a ser als tròpics, com a la seva formació, que acostuma a ser en masses d' aire marítimes tropicals. Per altra banda, anomenarem **sistema** a un conjunt específic de condicions climatològiques, que tenen en compte el moviment del vent a l' atmosfera.

Tenint aquests dos conceptes al cap podem definir un **cicló tropical** com un sistema tempestós caracteritzat per una baixa pressió al centre, seguit per una disposició en espiral de tempestes elèctriques. Els ciclons tropicals es creen gràcies a l' evaporació intensa de la superfície del mar, l' aire calent i humit s' inestabilitza, generant fortes corrents ascendents que donen lloc a una ràpida condensació del vapor d' aigua, que provoquen fortes precipitacions i vents huracanats que giren al voltant del centre del cicló. La característica que els diferencia de la resta de ciclons és que per a qualsevol altura atmosfèrica, el centre dels ciclons tropicals és més calent que els seus voltants.



Gràfic 1: Formació dels ciclons.

Depenent de la seva localització i la seva força, un cicló tropical pot classificar-se com un huracà o un tifó si els vents màxims a terra són d' almenys 33 m/s, una tempesta tropical si són almenys de 17 m/s, o una depressió tropical si són de menys de 17 m/s.

Per altra banda hi ha altres fenòmens meteorològics que es troben registrats en la nostra base de dades i seria interessant descriure'ls breument.

- Onada tropical: Àrea allargada de baixa pressió que causa nuvolositat i turmentes i que pot conduir a la formació de ciclons tropicals.
- Pertorbació tropical: Un sistema de vents ciclònics situats a les regions tropicals, de menor intensitat que un cicló tropical.
- Tempestes subtropicals: Tempestes localitzades a les zones tropicals o subtropicals, tenen la característica de ser menys intenses que les tempestes tropicals. Les tempestes subtropicals també poden conduir a la formació de ciclons tropicals.
- Tempestes extratropicals: Són sistemes de baixa pressió amb fronts freds, càlids i oclusos¹ associats. Per altra banda, aquestes tempestes presenten la seva velocitat màxima a uns 12 quilometres de terra a diferència de les tempestes tropicals. És curiós, però, que acostuma a passar que una tempesta tropical es transformi en una tempesta extratropical i a l' inrevés.

2.2 Manipulació de la base de dades

En aquest apartat explicarem, de manera general, tots els passos que van ser necessaris per a preparar la base de dades final amb la que treballarem.

El primer pas va ser obtenir quatre variables independents que ens indiquin l' hora, el dia, el mes i l' any en que es va originar la tempesta tropical. Per a aconseguir-ho, es van importar les dues bases de dades a Excel i en el procés d' importació es va indicar que es separés l' informació que contenia la variable en quatre de diferents, que evidentment, anomenem hora, dia, mes i any respectivament.

Un cop aconseguit això, es van importar les dues bases de dades a l' R, que com ja hem mencionat prèviament, serà el programa que utilitzarem per a dur a terme aquest treball. Tot seguit introduïm una nova variable a cada una de les bases de dades que s' anomenarà localització i ens indicarà si la tempesta tropical s' ha originat al Pacífic o a l' Atlàntic. Gràcies a aquest últim pas, podem unir les dues bases de dades en una de sola sense perdre la capacitat de poder diferenciar els ciclons segons la seva localització.

¹ Un front oclús es forma on un front càlid mòbil més lent és seguit per un front fred amb desplaçament més ràpid. El front fred atrapa al front càlid i l'empeny cap amunt. Els dos fronts continuen movent-se un darrera l'altre, i línia entre ells és el front oclús.

Tot seguit, es van escollir les variables que es creien rellevants per a l' estudi i es va obtenir una subbase de dades. Aquestes variables van ser: ID, nom, mes, dia, hora, any, velocitat màxima del vent, radi de la velocitat màxima del vent, radi de l' isòbar més proper, tipus i localització. Amb aquesta base de dades realitzarem l' anàlisi de les sèries temporals del número de ciclons anuals i mensuals, ja que per a realitzar-lo les dades faltants no són un problema i, d' aquesta manera no perdem informació.

A continuació es va realitzar un estudi de dades faltants, primerament de les variables de la base de dades obtinguda fins ara, i després d' aquestes mateixes variables però per cada un dels ciclons independentment, ja que, a l' ultima part d' aquest treball estudiarem, com ja hem comentat prèviament, tres índexs associats a l' intensitat dels ciclons i que definirem a partir de les variables de la base de dades, i no serà, per tant, adient, que un cicló presenti un percentatge molt elevat de dades faltants per a alguna de les seves variables.

Al realitzar l' anàlisi de dades faltants sobre les variables observades per cada cicló, veiem que en molts casos presenten un percentatge de missings massa elevat. El que es va fer llavors va ser eliminar tots aquells ciclons que presenten per a qualsevol de les seves variables un percentatge de dades faltants superior al 30%. Aquesta mesura va suposar passar a treballar amb 149 ciclons diferents, tenint, inicialment, 243.

Per altra banda, gràcies a l' estudi de dades faltats realitzat sobre la base de dades sencera, després de l' eliminació dels ciclons amb massa dades faltants, es va veure que les úniques variables que presentaven dades faltants eren les que mesuraven el radi de la velocitat màxima del vent, amb un 18.47%, i el radi de l' isòbar més proper, amb un 18.34%. En aquest cas, el percentatge de missings es va considerar acceptable i no es van eliminar més observacions.

Finalment es va procedir a estimar les dades faltants que encara presenta la nostra base de dades utilitzant la funció d' R *na.spline()*, que expliquem detalladament en el següent subapartat.

Per últim obtenim la velocitat màxima del vent, el radi de la velocitat màxima del vent i el radi de l' isòbar més proper, en les unitat del sistema internacional (metres i segons), ja que serà necessari tenir-ho en aquestes unitats per a calcular els índexs que s' exposen en l' apartat 4 d' aquest treball. Per fer-ho, multipliquem el valor actual de la velocitat màxima de vent per 0.514, ja que un knot equival a 0.514 m/s, i multipliquem el valor actual del radi de la velocitat màxima del vent i del radi de l' isòbar més proper per 1852, ja que una milla nàutica equival a 1852 metres.

Tots els passos anteriors ens porten a construir la base de dades que finalment utilitzarem per al nostre anàlisi, que consta de 11 variables i 8554 observacions.

Podem observar les primeres observacions que es troben a la nostra base de dades a la taula següent,

ID	Nom	Mes	Dia	Hora	Any	V. màx	R. V. màx.	R. isò.	Tipus	Loc.
EP0101	ADOLPH	5	25	18	2001	12.86	106607.6	860878.4	Sistema tropical	Pacífic
EP0101	ADOLPH	5	26	0	2001	15.43	111120	333360	Sistema tropical	Pacífic
EP0101	ADOLPH	5	26	6	2001	15.43	111120	166680	Sistema tropical	Pacífic
EP0101	ADOLPH	5	26	12	2001	15.43	111120	166680	Sistema tropical	Pacífic
EP0101	ADOLPH	5	26	18	2001	18	111120	166680	Sistema tropical	Pacífic
EP0101	ADOLPH	5	27	0	2001	20.57	74080	324100	Sistema tropical	Pacífic

Taula 1: Visualització de les primeres observacions de la base de dades, on V. màx, fa referència a la velocitat màxima del vent, R. V. màx, al radi de la velocitat màxima del vent, R. isò al radi de l' isòbar més proper i Loc a la localització.

2.3 Estimació de dades faltants

La funció *na.spline* que hem utilitzat per a estimar els valors faltants de la nostra base de dades, substitueix cada dada faltant per un valor obtingut amb una funció cúbica ajustada a les nostres dades. El que hem fet ha estat aplicar-la, a cada cicló, a les dues variables que presentaven dades faltants, radi de la velocitat màxima del vent i radi de l' isòbar més proper.

Aprofundint més en el seu funcionament, cal tenir en compte que, concretament, el que fa aquesta funció, és una **interpolació cúbica amb splines**. En matemàtiques, el concepte spline fa referència a una funció definida per parts i que posseeix un alt grau de suavitat en els llocs on les parts s'ajunten, i al dir que utilitzem una interpolació cúbica, el que especifiquem és que aquesta funció serà un polinomi de tercer grau.

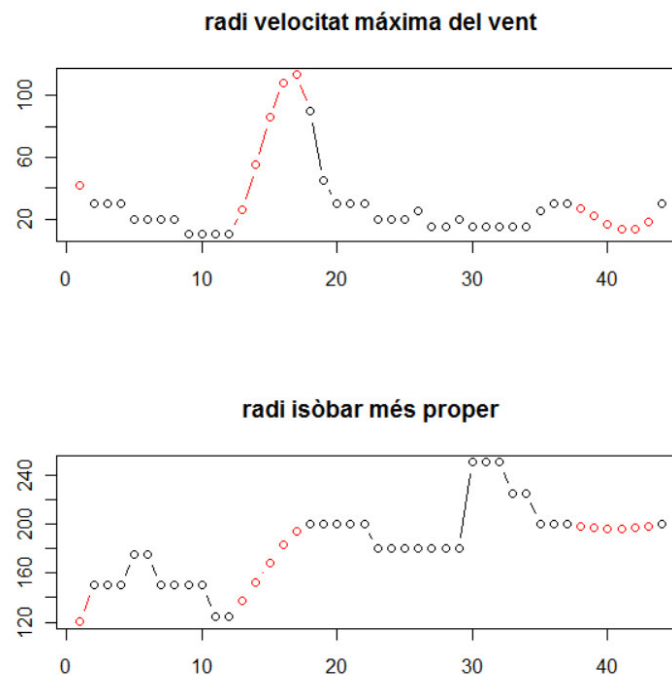
Més detalladament, podem dir que, donada una funció f definida a l'interval $[a, b]$ i amb els següents trossos,

$$a = x_0 < x_1 < x_2 < \dots < x_n = b$$

Una interpolació cúbica amb splines, S , per a f , presenta les següents característiques,

- S_j és un polinomi cúbic definit a l'interval $[x_j, x_{j+1}]$ per a j des de 0 a $n - 1$.
- $S(x_j) = f(x_j)$ per a j de 0 a n .
- $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$ per a j de 0 a $n - 2$.
- $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$ per a j de 0 a $n - 2$.
- $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$ per a j de 0 a $n - 2$.
- $S''(x_0) = S''(x_n) = 0$
- $S'(x_0) = f'(x_0)$ i $S'(x_n) = f'(x_n)$

Gràcies a les característiques anteriors, i tenint en compte els valors que coneixem, podem calcular els coeficients dels S_j polinomis com a solució d' un sistema d' equacions². Per a acabar d' entendre el seu funcionament, podem visualitzar el següent gràfic que ens mostra en vermell l' estimació obtinguda per a les variables que presentaven dades faltants per al cicló Adrian, que va tenir lloc al maig de 2005 al Pacífic.



Gràfic 2: Exemple de l' aplicació de la funció na.spline en el Cicló Adrian, al maig del 2005. Els punts vermells fan referència a l' estimació obtinguda amb la funció.

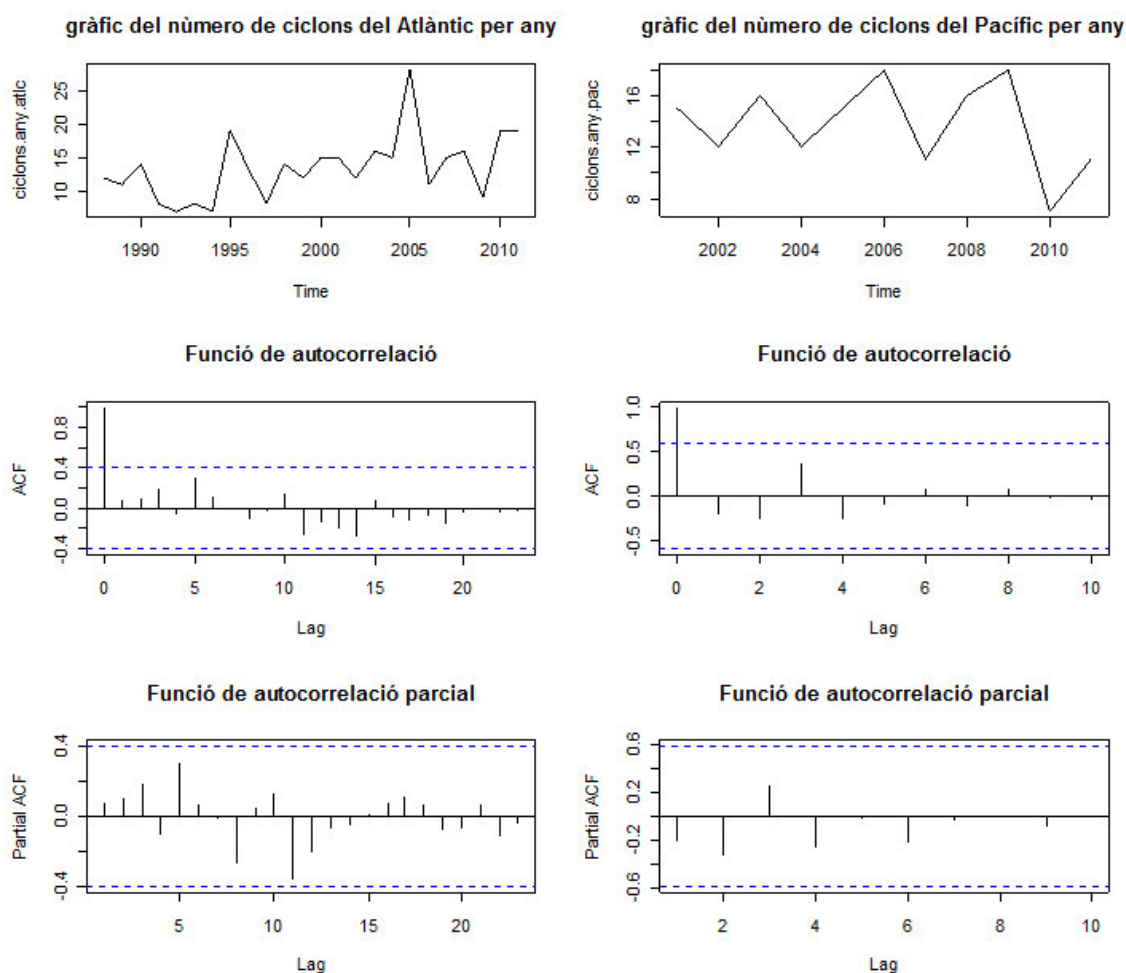
² Per a ampliar l' informació sobre l' interpolació cúbica amb splines es pot consultar el següent document: <http://banach.millersville.edu/~BobBuchanan/math375/CubicSpline/main.pdf>

3. Anàlisi de sèries temporals

En aquest apartat procedirem a estudiar les sèries temporals del número de ciclons anuals i mensuals diferenciant segons s' hagin produït al Pacífic o a l' Atlàntic. És clar, doncs, que com hem comentat a l' apartat anterior, per a realitzar aquest anàlisi podem utilitzar la base de dades prèvia al tractament de dades faltants, ja que aquestes no ens afecten a l' hora de veure el número de ciclons que tenim registrats i, d' aquesta manera, no perdrem informació.

3.1 Estudi de les sèries anuals

Podem començar estudiant les sèries anuals del número de ciclons, mitjançant els següents gràfics (Gràfic 3).



Gràfic 3: Sèrie temporal del número de ciclons anual registrats al Atlàntic i al Pacífic, gràfic de la funció d'autocorrelació i d'autocorrelació parcial per a cada sèrie.

Podem veure que cap de les dues sèries presenta una estructura clara, és a dir, sembla que les observacions siguin incorrelades. A més, si ens fixem en els gràfics d' autocorrelació i d' autocorrelació parcial, ens porten a pensar que aquestes sèries podrien definir-se com a soroll blanc.

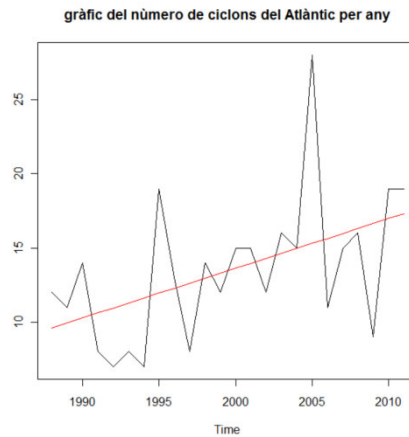
Per a poder confirmar o rebutjar les nostres sospites, utilitzem el test de Box Ljung que contrasta l' hipòtesi nul·la de que les dades són independentment distribuïdes, donat un determinat lag temporal. Els resultats obtinguts, per a les dues sèries i per a qualsevol lag de temps possible, ens indiquen que no tenim prou evidències com per a rebutjar l' hipòtesi nul·la³ de que les sèries efectivament es poden definir com a soroll blanc. És a dir, considerem que les sèries del número de ciclons anuals tant per l' Atlàntic com pel Pacífic, consisteixen en una successió d' observacions aleatòries centrades, amb la mateixa variància i incorrelades dos a dos, que anomenem *soroll blanc*, i presenten les següents propietats:

1. $E[Z_t] = 0 \quad \forall t \in \mathbb{Z}$
2. $E[Z_t^2] = \sigma^2 \quad \forall t \in \mathbb{Z}$
3. $\text{Cov}(Z_s, Z_t) = 0, \quad \forall t \neq s$

També s' estudia la tendència lineal d' aquestes dues sèries a partir d' un model de regressió lineal simple que intenta explicar el número de ciclons registrats anualment, variable resposta, a través del any en el que s'han observat, variable explicativa, ja que, visualment, la sèrie del número de ciclons anuals sembla presentar una tendència lineal. Gràcies a aquest model podem veure que en el cas del Atlàntic, el pendent lineal resulta significatiu, amb un p-valor de 0.015, el que ens fa pensar que si que es pot considerar certa linealitat, a més, com que l' estimació d' aquest pendent és positiva, concretament 0.3, pensarem que el número de ciclons anuals del Atlàntic presenta una tendència lineal creixent, tal com podem observar en la seva gràfica. En canvi, al realitzar el mateix model per a la sèrie de ciclons del Pacífic, veiem que el pendent no resulta significatiu, amb un p-valor igual a 0.43, el que ens fa pensar que la sèrie no presenta una clara tendència lineal.

A continuació, podem observar el gràfic de l' ajust de la tendència lineal de la sèries del número de ciclons anuals a l' Atlàntic,

³ Podem trobar la taula de p-valors obtinguts en cada cas als annexos, apartat A1.



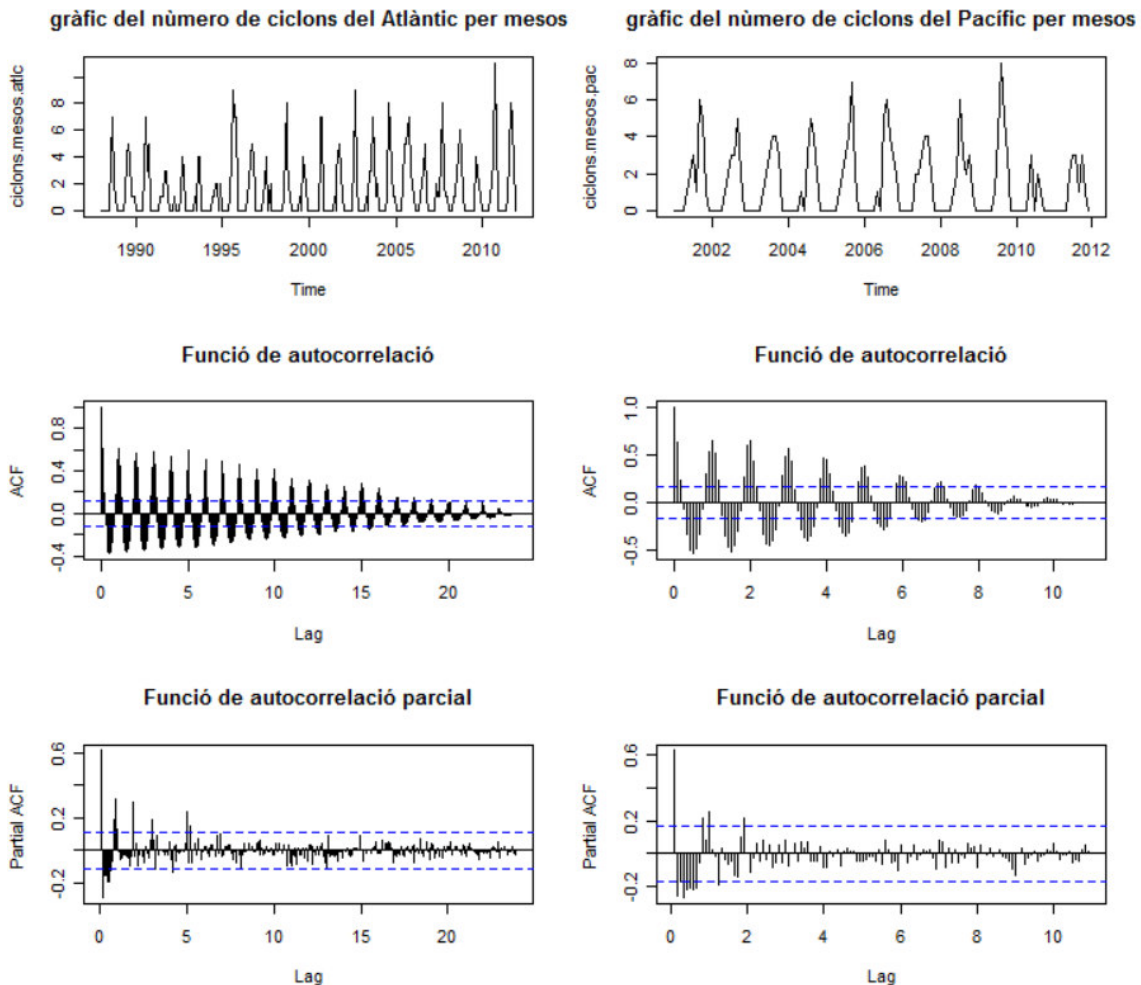
Gràfic 4: Ajust de la tendència lineal de la sèrie de número de ciclons anuals a l' Atlàntic

Per últim, com que en el cas del Atlàntic obtenim que presenta una tendència lineal, estadísticament significativa, estudiem la sèrie sense tendència, és a dir, els residus que obtenim sota el model de regressió lineal. En aquest cas, únicament mirarem que siguin incorrelats, mitjançant el test de Box Ljung, i obtenim que efectivament així és, per a qualsevol lag de temps. Els resultats es podem veure a l' annex, apartat A2.

3.2 Estudi de les sèries mensuals

Un cop finalitzat l' anàlisi de les sèries anuals, passem a analitzar les sèries mensuals a través de la visualització del seu gràfic temporal, la funció d'autocorrelació i la d'autocorrelació parcial (Gràfic 4). Aquests gràfics són un indicatiu de que aquestes sèries podrien ser considerades SARIMA⁴ amb periodicitat 12, ja que són sèries mensuals. D' entrada, el fet de que les sèries siguin mensuals ja ens fa pensar en un procés SARIMA, però, els gràfics d'autocorrelació i d'autocorrelació parcial que observem en són un indicatiu encara més clar, ja que presenten l' estructura típica d'un procés SARIMA.

⁴ Si d i D són dos enters no negatius, X_t és defineix com un procés $SARIMA(p,d,q) \times (P,D,Q)_s$ amb periodicitat s si, $Y_t = (1 - B)^d(1 - B^s)^D X_t$ és un procés causal ARMA. A més, Direm que una sèrie estacionària $\{Z_t, t \in \mathbb{Z}\}$ és una sèrie autoregressiva amb mitjana mòbil d'ordres p i q respectivament o $ARMA(p,q)$, si compleix una equació, $X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$ on $\{Z_t, t \in \mathbb{Z}\}$ és un soroll blanc i els polinomis $\Phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p$ i $\Theta(x) = 1 + \theta_1 x + \dots + \theta_q x^q$ tenen totes les arrels fora del cercle unitat i no tenen arrels comuns.



Gràfic 5: Sèrie temporal del número de ciclons mensuals registrats al Atlàntic i al Pacífic, gràfic de la funció d'autocorrelació i d'autocorrelació parcial per a cada sèrie.

Per altra banda podem veure que, la funció d'autocorrelació, que decreix molt lentament, no entra dins de l'interval de confiança per a la correlació igual a zero fins a un lag de temps molt elevat. Per a evitar això, diferenciem una vegada la sèrie inicial, com podem observar als nous gràfics d'autocorrelació i d'autocorrelació parcial als annexos, apartat 3, d'aquesta manera la funció d'autocorrelació entra abans a l'interval de confiança de la correlació igual a zero. Al mateix apartat del annexos podem trobar també la sèrie sense estacionalitat, amb els seus gràfics d'autocorrelació i d'autocorrelació parcial. Per aquesta raó, els paràmetres d i D del procés SARIMA els agafem ser iguals a 1, ja que aquests indiquen el número de vegades que ha estat necessari diferenciar la sèrie inicial.

Tenint això clar, passem a estudiar la sèrie del número de ciclons mensuals del Atlàntic i del Pacífic. El procediment habitual per a estudiar aquest tipus de sèries es basa en establir totes les combinacions possibles dels valors p , q , P , i Q per a

valors iguals a 0 o a 1⁵. Un cop establerts aquests models s'escull el millor seguint el criteri del mínim AIC⁶ i es realitza un anàlisi dels seus residus per a poder decidir si és un model vàlid per a aplicar a les nostres dades. Al establir tots els models possibles, veiem que en el cas de la sèrie del Atlàntic, el mínim AIC obtingut s'observa sota la hipòtesis de que la sèrie és un procés SARIMA(1,1,1) x (0,1,1)₁₂ i amb molt poca diferència per a un SARIMA(1,1,1) x (1,1,1)₁₂, en canvi, per a la sèrie del Pacífic, el millor model sembla ser el que s'obté per a un procés SARIMA(0,1,1) x (1,1,1)₁₂.

Finalment, realitzem l'anàlisi dels residus per veure si aquests models són realment vàlids per a aplicar-los a les nostres dades⁷ i per decidir quin dels dos models estudiats és millor per a modelar la sèrie del número de ciclons de l'Atlàntic. Recordem que els residus s'obtenen restant a les dades originals les estimacions obtingudes amb un model determinat, és a dir, suposarem que les nostres dades segueixen un model determinat amb el que obtindrem unes estimacions i, per tant, els residus. Recordem també que els residus es podem interpretar com una estimació del valor i-èssim del soroll blanc que defineix la sèrie que conformen les nostres dades i per tant esperarem que siguin incorrelats, encara que com es tracta d'una estimació, en tot cas seran aproximadament incorrelats. Tot i que no és una hipòtesi necessària, estudiarem també la normalitat dels errors obtinguts en cada cas.

Comprovem les nostres hipòtesis mitjançant tests estadístics i, concretament, podem dir que no són normals, utilitzant el test de Shapiro-Wilk, obtenint un p-valor de $0.6 \cdot 10^{-16}$ en el cas del Atlàntic sota un model SARIMA(1,1,1) x (1,1,1)₁₂ i $2.5 \cdot 10^{-16}$ sota un model SARIMA(1,1,1) x (0,1,1)₁₂, pel que fa al Pacífic obtenim un p-valor igual a $1.8 \cdot 10^{-9}$, per altra banda, el que si que podem dir és que són incorrelats segons el test de Box Ljung, amb un p-valor de 0.97 per a l'Atlàntic (en tots dos casos) i 0.12 per al Pacífic⁸.

Per tant, tenint en compte els resultats anteriors, és a dir, que podem confirmar l'hipòtesis d'incorrelació en tots dos casos, i ja que no s'observem grans diferències entre els dos models seleccionats per a l'Atlàntic, proposem que la sèrie del número de ciclons mensuals de l'Atlàntic segueix un procés SARIMA(1,1,1) x (0,1,1)₁₂ mentre que la del Pacífic segueix un procés SARIMA(0,1,1) x (1,1,1)₁₂.

Un altre estudi de caire més descriptiu relacionat amb les sèries del número de ciclons mensuals, vol veure si hi ha alguns mesos que presentin un número de ciclons major que altres. Això ho podem veure a través de la següent taula resum,

⁵ Els valors dels paràmetres d'aquests processos acostumen a ser inferiors a 3, però és comú utilitzar valors iguals a 1 i 0 per a realitzar l'estudi de la sèrie, tot i que es podrien realitzar totes les combinacions possibles dels valors entre 0 i 3.

⁶ Es pot trobar una taula amb tots els AIC obtinguts als annexos, apartat A4.

⁷ Es poden trobar alguns gràfics relacionats amb l'anàlisi dels residus als annexos, apartat A5 per a l'Atlàntic i A6 per al Pacífic.

⁸ El p-valor del test de Box Ljung es dona per a un lag igual a 4 en tots dos casos, però s'ha comprovat que el p-valor obtingut no és inferior a 0.05 en lags de fins a 20.

	Atlàntic												Pacífic											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
1988	0	0	0	0	0	0	0	4	7	2	1	0												
1989	0	0	0	0	0	1	4	5	4	1	1	1												
1990	0	0	0	0	0	0	3	7	3	5	0	0												
1991	0	0	0	0	0	1	1	1	3	3	1	0												
1992	0	0	0	1	0	0	0	1	4	3	0	0												
1993	0	0	0	0	0	1	0	4	4	0	0	0												
1994	0	0	0	0	0	1	1	2	2	0	2	0												
1995	0	0	0	0	0	1	4	9	7	7	1	0												
1996	0	0	0	0	0	1	2	4	5	4	1	0												
1997	0	0	0	0	1	2	4	0	1	2	0	0												
1998	0	0	0	0	0	0	1	5	8	4	2	1												
1999	0	0	0	0	0	1	0	4	4	3	2	0												
2000	0	0	0	0	0	0	0	4	7	7	0	0												
2001	0	0	0	0	0	1	0	3	4	5	3	1	0	0	0	0	1	2	3	1	6	5	1	0
2002	0	0	0	0	0	0	1	3	9	2	0	0	0	0	0	0	1	2	3	3	5	2	0	0
2003	0	0	0	1	0	1	3	3	7	4	1	2	0	0	0	0	1	2	3	4	4	3	0	0
2004	0	0	0	0	0	0	1	8	6	3	1	1	0	0	0	0	1	0	3	5	4	2	0	0
2005	0	0	0	0	0	2	5	5	6	7	3	2	0	0	0	0	1	2	3	4	7	2	0	0
2006	1	0	0	0	0	1	2	3	5	1	0	0	0	0	0	0	1	0	5	6	4	3	2	0
2007	0	0	0	0	2	1	1	4	8	2	1	1	0	0	0	0	2	2	3	4	4	2	0	0
2008	0	0	0	0	1	1	3	4	6	4	1	0	0	0	0	0	1	2	6	3	2	3	1	0
2009	0	0	0	0	0	0	0	4	3	2	1	0	0	0	0	0	0	1	4	8	5	3	0	0
2010	0	0	0	0	0	1	2	4	11	5	1	0	0	0	0	0	1	3	0	2	1	0	0	0
2011	0	0	0	0	0	1	4	8	7	3	1	0	0	0	0	0	0	2	3	3	1	3	1	0

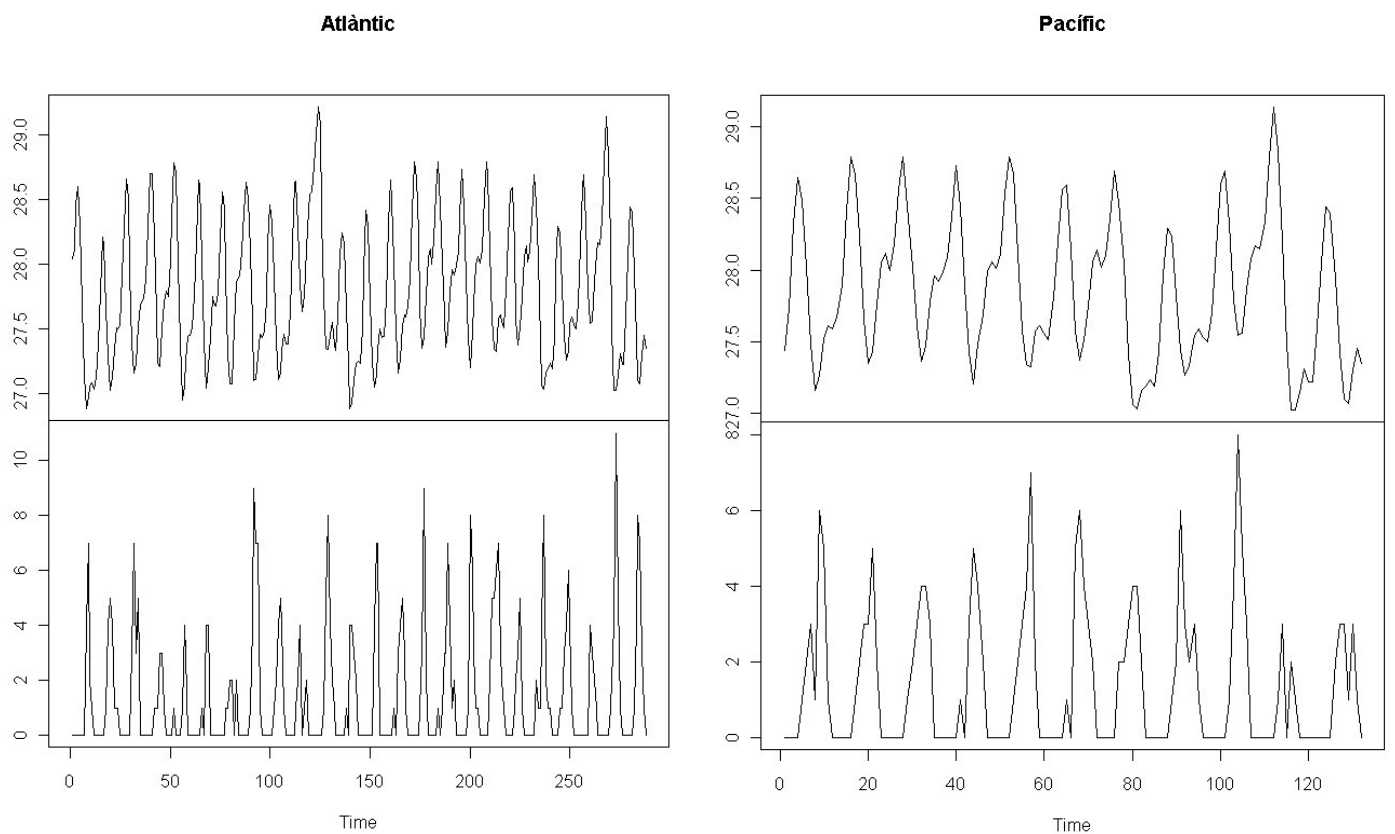
Taula 2: Número de ciclons mensuals per als diferents anys registrats a l' Atlàntic i Pacífic.

Gràcies a la taula anterior podem determinar que els mesos en els que es registren un major número de ciclons mensuals normalment són Agost i Setembre. Això té sentit tenint en compte que els ciclons es formen com a conseqüència de la evaporació del aigua i, per tant, serà necessari una temperatura elevada. Això ens fa pensar en un nou anàlisi descriptiu, introduint la temperatura com a factor per a explicar el número de ciclons obtinguts mensualment.

Per a realitzar-lo el que farem serà observar conjuntament la sèrie mensual de ciclons, a l' Atlàntic i al Pacífic, amb la sèrie temporal de les temperatures mensuals, en graus Centígrads, de la superfície del mar obtingudes als Tròpics (10°N-10°S, 0-360°). Aquesta informació l' obtenim gràcies al *National Oceanographic and Atmospheric Administration (NOAA)*, i concretament es podem obtenir a través de la següent pàgina,

<http://www.cpc.ncep.noaa.gov/data/indices/sstoi.atl.indices>

On la variable que ens interessa, la que guarda l' informació de la temperatura mensual als Tròpics, és TROP.



Gràfic 6: Sèrie temporal del número de ciclons mensuals registrats al Atlàntic i al Pacífic (gràfics de sota) i sèrie de la temperatura mensual als Tròpics (a dalt).

En aquests gràfics podem observar el que ja sospitàvem després de analitzar la taula 2, que el número de ciclons augmenta a conseqüència d'un augment en la temperatura. Fixem-nos, concretament, com després de cada pic a la taula de a dalt, la que ens mostra la temperatura, obtenim un pic a la de sota, la del número de ciclons mensuals.

Aquest anàlisi es podria aprofundir molt més, donant lloc a un nou treball. En aquest cas, però, no és l'objectiu d'aquest estudi i únicament es realitza un anàlisi superficial.

4. Índexs PDI, IEI i OEI

En meteorologia s' utilitzen tres índex diferents per a definir l' intensitat dels ciclons. Aquests es defineixen de la següent manera,

- Índex **PDI** d' un cicló associat a dos instants de temps t_0 i t_n

$$\int_{t_0}^{t_n} v_{max}^3 dt,$$

on v_{max} és la velocitat màxima del vent en m/s. A la pràctica, però, aproximarem la integral per la suma discreta lligada a la partició $t_0 \leq t_1 \leq \dots \leq t_n$,

$$\sum_{i=0}^n v_{max}^3 (t_{i+1} - t_i)$$

- Índex **IEI** d' un cicló associat a dos instants de temps t_0 i t_n

$$\int_{t_0}^{t_n} v_{max}^3 R_{max}^2 dt,$$

on R_{max} és el radi de màxima velocitat del vent en metres. En aquest cas aproximarem la integral per la suma discreta lligada a la partició $t_0 \leq t_1 \leq \dots \leq t_n$,

$$\sum_{i=0}^n v_{max}^3 R_{max}^2 (t_{i+1} - t_i)$$

- Índex **OEI** d' un cicló originat entre dos instants de temps t_0 i t_n

$$\int_{t_0}^{t_n} v_{max}^3 R_{iso}^2 dt,$$

on R_{iso} és el radi de l' isòbar més proper en metres. En aquest cas aproximarem la integral per la suma discreta lligada a la partició $t_0 \leq t_1 \leq \dots \leq t_n$,

$$\sum_{i=0}^n v_{max}^3 R_{iso}^2 (t_{i+1} - t_i)$$

I, per a tots els casos, donat que la base de dades dona les mesures de les variables dels ciclons cada sis hores, les diferències de temps seran sempre de sis hores, el que equival a 21600 segons.

Com hem fet durant tot el treball, diferenciem segons si els ciclons s' han donat a l' Atlàntic o al Pacífic, ja que s' esperen observar diferències.

4.1 Estudi de la relació entre índexs

En primer lloc estudiarem la relació entre aquests índexs mitjançant la seva matriu de correlacions, gràfics de dispersió dos a dos i l'estimació de models de regressió lineal dos a dos, tal com podem veure a continuació.

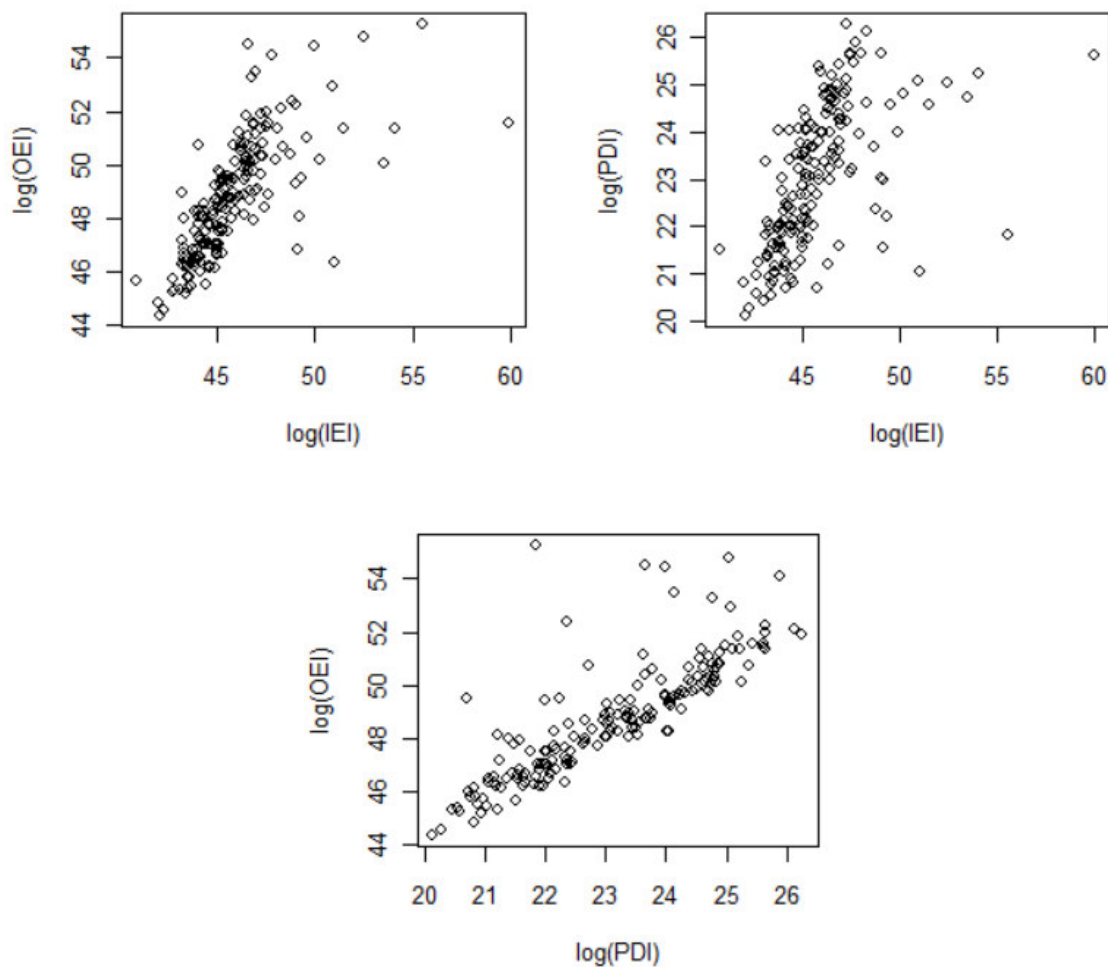
Comencem observant els resultats obtinguts per als índexs del Atlàntic. El primer que es calcula es la matriu de correlacions per als índexs i a l'observar que no presenten una correlació molt elevada, es passa a calcular la matriu de correlacions entre el logaritme dels índexs, ja que es pensa que també podria estar correlacionat. En aquest últim cas podem observar com les correlacions augmenten, el que ens indica clarament que els logaritmes del índexs es troben, d'alguna manera, correlacionats.

	PDI	IEI	OEI
PDI	1	0.009	0.2
IEI	0.009	1	0.14
OEI	0.2	0.14	1

	Log(PDI)	Log(IEI)	Log(OEI)
Log(PDI)	1	0.69	0.57
Log(IEI)	0.69	1	0.83
Log(OEI)	0.57	0.83	1

Taula 3: Matriu de correlacions entre els índexs del Atlàntic i entre els logaritmes dels índexs.

D'aquesta manera també veiem que la relació més forta es troba entre el logaritme de l'índex IEI i el de l'índex OEI. Per a aprofundir en aquest anàlisi realitzem, com hem comentat prèviament, diagrames de dispersió dos a dos. Aquests diagrames els realitzarem directament entre els logaritmes dels índexs ja que és on s'observa una correlació més elevada (Gràfic 7). En aquest gràfic ja s'observa certa tendència lineal que queda confirmada al aplicar els models de regressió simple. Concretament podem dir que el pendent entre el logaritme de l'índex IEI i l'OEI és de 0.77, per altra banda, el pendent entre el logaritme de l'índex IEI i el PDI és de 0.93 i finalment, el pendent entre el logaritme de l'índex PDI i l'OEI és de 0.56, amb un p-valor menor a $2 \cdot 10^{-16}$ en tots els casos.



Gràfic 7: Diagrames de dispersió dos a dos entre els logaritmes dels índexs del Atlàntic.

Un cop finalitzat l'anàlisi per als índexs dels ciclons de l'Atlàntic, passem a repetir el mateix procediment, ara per als índexs dels ciclons del Pacífic. Com abans, comencem calculant les matrius de correlacions, entre els índexs i entre els seus logaritmes.

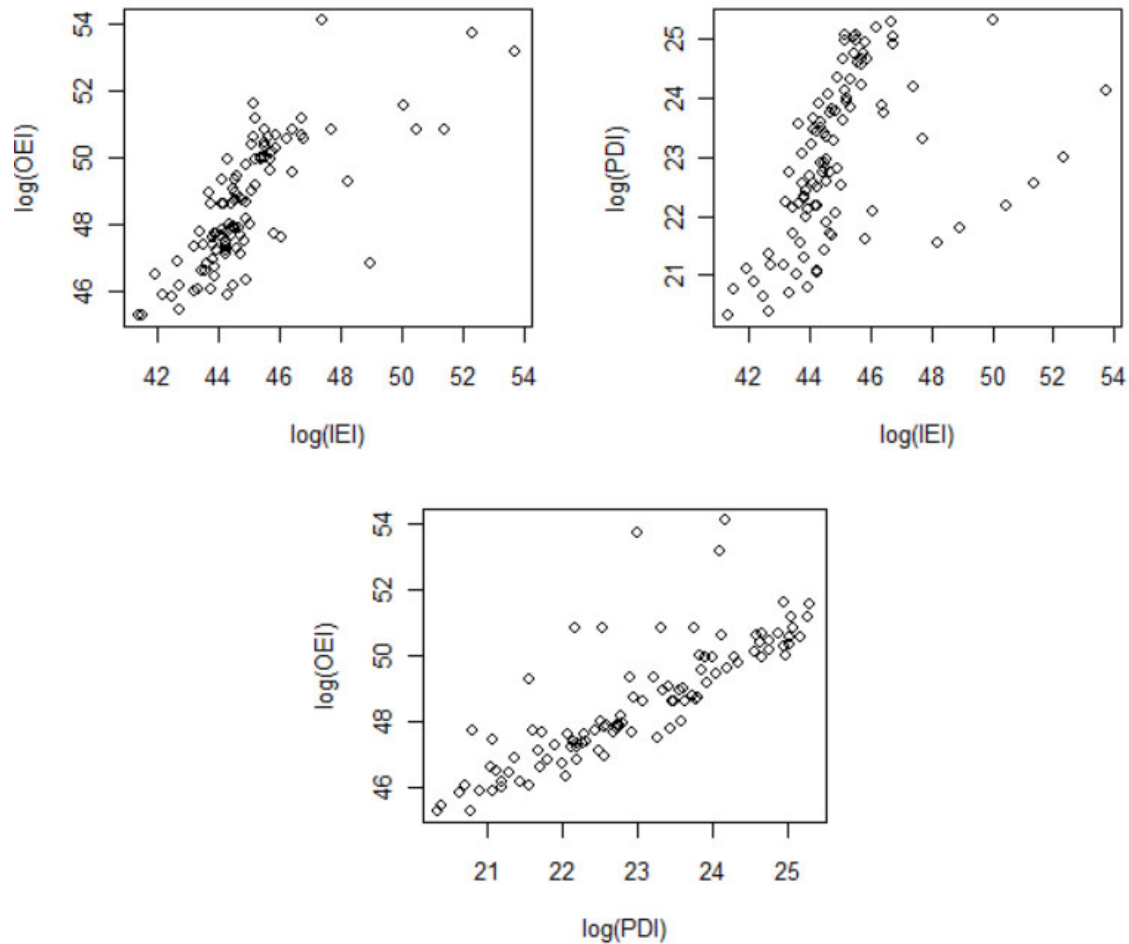
	PDI	IEI	OEI
PDI	1	0.4	0.02
IEI	0.4	1	0.12
OEI	0.02	0.12	1

	Log(PDI)	Log(IEI)	Log(OEI)
Log(PDI)	1	0.74	0.43
Log(IEI)	0.74	1	0.83
Log(OEI)	0.43	0.83	1

Taula 4: Matriu de correlacions entre els índexs del Pacífic i entre els logaritmes dels índexs.

En aquest cas també s'observa una correlació més elevada entre els logaritmes dels índexs i, concretament, tornem a obtenir que els logaritmes dels índexs que estan més correlacionats són els dels índexs OEI i IEI. Seguint els mateixos passos que per a l'anàlisi dels índexs de l'Atlàntic, ara visualitzarem els diagrames de dispersió dos a dos (Gràfic 8), on observem una tendència lineal que gràcies als

models de regressió lineal podem quantificar dient que el pendent entre el logaritme de l' índex IEI i l' OEI és de 0.79, per altra banda, el pendent entre el logaritme de l' índex IEI i el PDI és de 0.64 i finalment, el pendent entre el logaritme de l' índex PDI i l' OEI és de 0.59, amb un p-valor menor a $2 \cdot 10^{-16}$ en tots els casos.



Gràfic 8: Diagrames de dispersió dos a dos entre els logaritmes dels índexs del Pacífic.

Sembla interessant notar que els pendents obtinguts en el cas del Atlàntic i el Pacífic són molt semblants, llevat del pendent que es dona entre el logaritme dels índexs IEI i PDI que resulta ser bastant inferior en el cas del Pacífic.

4.2 Estimació de la funció de densitat

Un cop finalitzat aquest petit anàlisi de relació entre índexs, el que farem serà estimar la seva funció de densitat. Per a fer-ho, utilitzarem la funció *hist()* d' R, ja que una de les seves sortides és la densitat estimada. El que fa aquesta funció per a estimar la densitat és, en el cas en el que la diferència dels valors en els que mirem la freqüència de les dades (breaks) sigui en tots els casos igual a 1, estima la funció de densitat com la freqüència relativa, és a dir, el recompte d' observacions dins un cert interval dividit entre el total d' observacions, i en general, satisfà la següent restricció,

$$\sum_i \hat{f}(x_i)(b_{i+1} - b_i) = 1$$

On $\hat{f}(x_i)$ fa referència a l' estimació de la funció de densitat per a un determinat valor i b_i és el break i-èssim.

D' aquesta manera per al càlcul de la densitat l' únic que haurem de determinar seran els breaks que utilitzarem. En aquest cas el que farem servir serà un **logarithmic binning** de manera que obtenim els diferents breaks multiplicant el break anterior per una constant, així, establim com a primer break un número inferior al mínim observat en el determinat índex i li multipliquem una certa constant fins a obtenir un número més elevat que el màxim de l' índex⁹.

Per altra banda, **La llei de Potències** estipula que les freqüències d' una variable aleatòria decreixen segons augmenta la variable aleatòria, de manera que esperem que la funció de densitat sigui definida per la següent igualtat, sempre que x, en el nostre cas els índexs, sigui més gran que un determinat valor A positiu,

$$f(x) \approx \frac{c}{x^\alpha} 1_{[A, +\infty]} ,$$

on c és una constant que queda determinada a partir dels dos paràmetres, positius, A i α que defineixen la llei de potències i α és el que es coneix com l' exponent de la potència. Podem veure, a continuació, un exemple gràfic de la llei de potències,



Gràfic 9: Exemple gràfic de la llei de potències. A la dreta es troba la cua llarga amb poques observacions per a cada valor, a l' esquerra podem trobar molts elements en els valors que podem considerar freqüents.

⁹ Per a poder veure el procediment aplicat a cada cas anar als annexos, apartat A7, on trobem la sintaxis d' R necessària.

Gràcies a l' exponent de la potència podem veure que,

- Si $2 \geq \alpha > 1$, l'esperança de la variable aleatòria és infinita.
- Si $3 \geq \alpha > 2$, la variable aleatòria tindrà esperança però no moment de segon ordre.
- Si $n \geq \alpha > n - 1$, la variable aleatòria tindrà moments finits fins l' orde $n-2$.

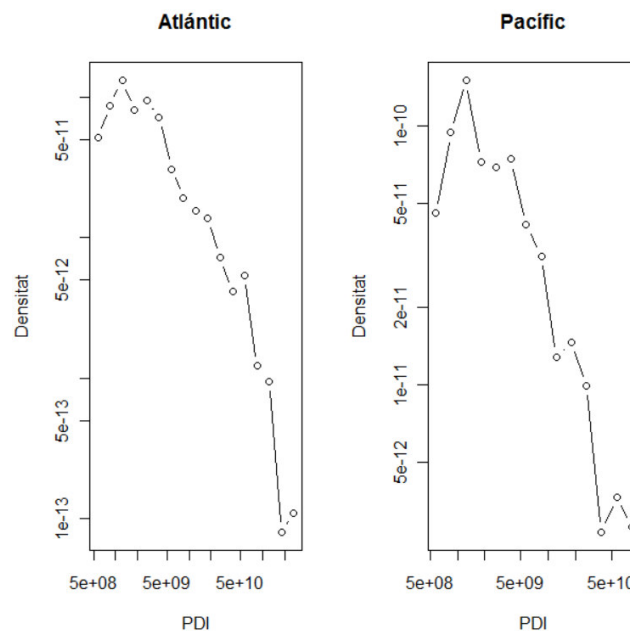
A més, evidentment, al aplicar logaritmes,

$$\ln(f(x)) = \ln C - \alpha \ln x$$

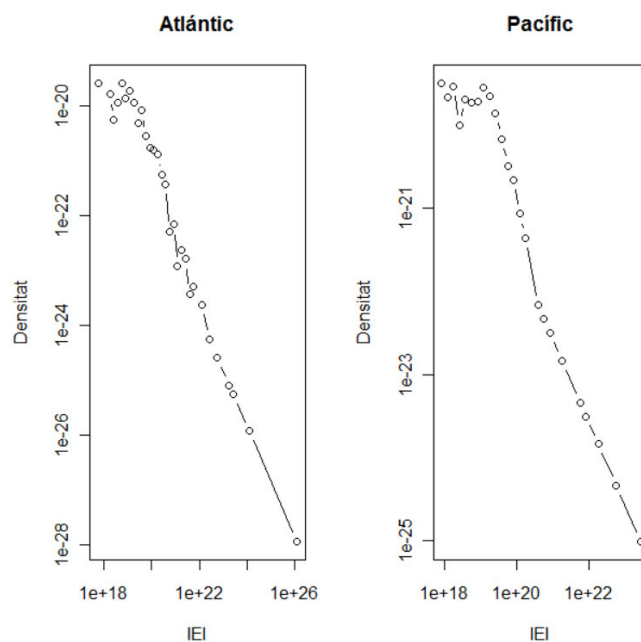
I per tant, en escala logarítmica obtindrem una recta.

Amb això clar, és important considerar un logarithmic binning exponencial, ja que segons la llei de potències esperem que la funció de densitat tingui una forma exponencial, amb el que aplicant logaritmes, els breaks seran equidistants i la funció de densitat passarà a tenir tendència lineal. En aquest cas, la constant que utilitzarem per a obtenir els breaks en cada cas serà $10^{\frac{1}{\alpha}}$.

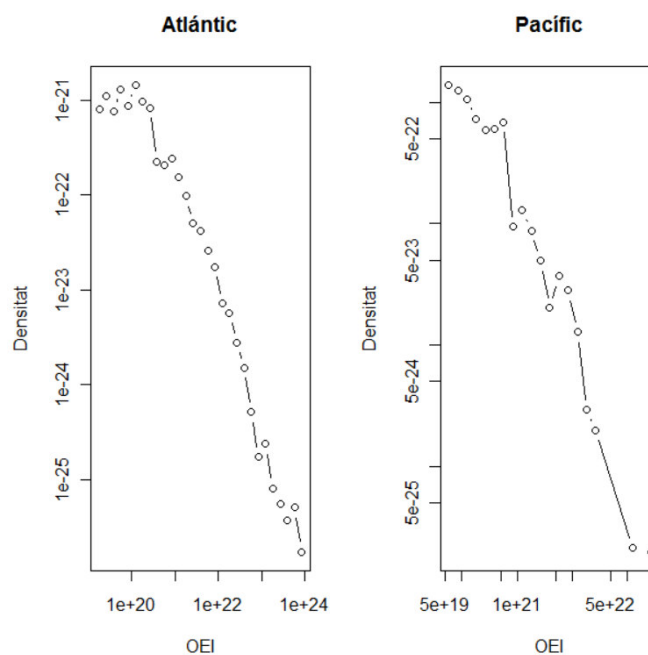
Així, sabent els breaks a aplicar, construïm l' histograma i n' extraiem l' estimació de la densitat. A continuació podem observar els gràfic d' aquesta estimació, utilitzant com a valor x, no els breaks, si no els punts centrals entre un break i el següent.



Gràfic 10: Estimació de la funció de densitat per a l' índex PDI (aplicant logaritmes).



Gràfic 11: Estimatció de la funció de densitat per a l'índex IEI (aplicant logaritmes).



Gràfic 12: Estimatció de la funció de densitat per a l'índex OEI (aplicant logaritmes).

Com podem observar als gràfics anteriors, les primeres observacions no s'adapten bé a la tendència lineal que esperem obtenir sota la llei de potències, i per tant, aquests primers valors no els tindrem en compte en apartats posteriors, considerant el valor a partir del qual es pot pensar que els índexs segueixen una llei de potències com el paràmetre A exposat anteriorment.

4.3 Estimació de l' exponent de potència

En tots els gràfics anteriors podem observar com, a partir d' un cert valor, l' estimació de la funció de densitat podria ajustar-se bé a través d' una recta, el pendent de la qual és l' exponent de la potència α suposant que la funció de densitat segueix la llei de potències, i degut a que apliquem logaritmes per tal de obtenir-la en forma lineal, es veuen les següents igualtats,

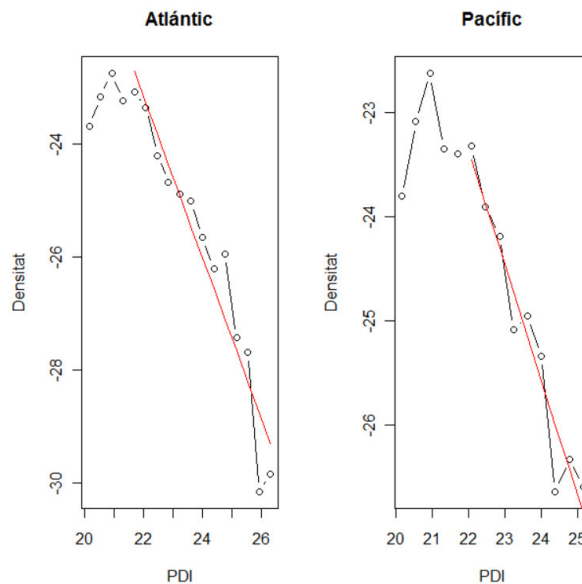
$$\ln(f(x)) = \ln C - \alpha \ln x$$

El que, al realitzar la regressió lineal equival a dir que,

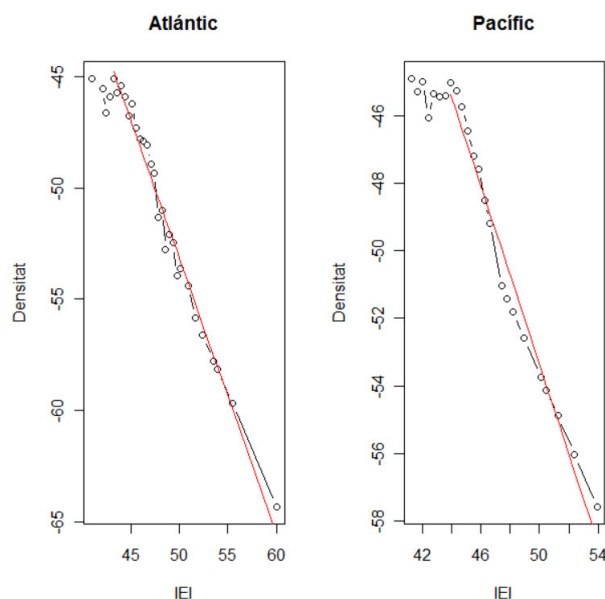
$$C = e^{\widehat{\beta}_0}$$

$$\alpha = -\widehat{\beta}_1$$

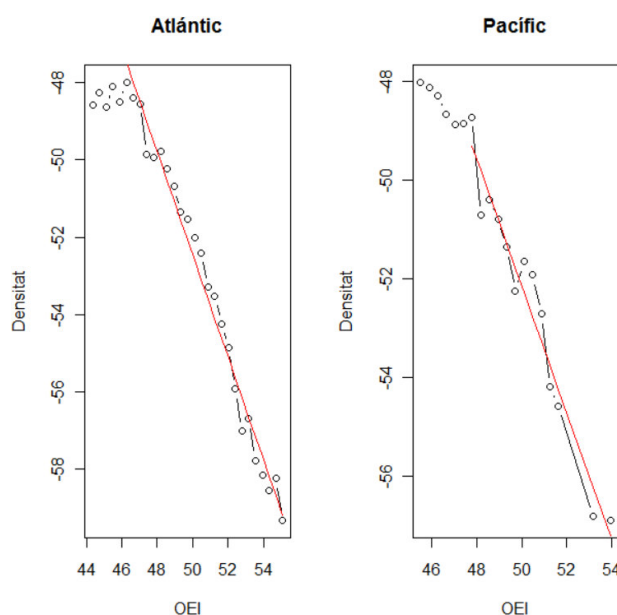
El següent pas lògic serà, doncs, estimar aquest paràmetre α per als tres índexs del Atlàntic i del Pacífic. Com que es tracta d' estimar un pendent, el que farem serà utilitzar models de regressió lineal on la variable resposta d' interès serà el logaritme de la densitat estimada i la variable explicativa serà el logaritme dels punts centrals entre dos determinats breaks, tenint en compte que no introduïrem en aquests models les primeres observacions ja que fan desviar la tendència lineal de les dades. Les estimacions obtingudes es podem trobar en la taula 5. A continuació es poden observar els gràfics que mostren l' ajust de la recta estimada mitjançant la regressió lineal sobre cada una de les funcions de densitat estimades,



Gràfic 13: Ajust de la recta de regressió sobre la funció de densitat per a l' índex PDI (aplicant logaritmes).



Gràfic 14: Ajust de la recta de regressió sobre la funció de densitat per a l'índex IEI (aplicant logaritmes).



Gràfic 15: Ajust de la recta de regressió sobre la funció de densitat per a l'índex OEI (aplicant logaritmes).

Per a finalitzar aquest subapartat, calcularem intervals de confiança per als diferents exponents de potències. La manera més ràpida d'obtenir un interval de confiança per al paràmetre α , pendent de les rectes anteriors, donat que hem utilitzat models de regressió lineal, seria utilitzar l'error estàndard estimat del coeficient $\widehat{\beta}_1$, però si féssim això estaríem sota la suposició de normalitat, cosa que no és gens evident. Per tant, el que farem, serà utilitzar els mètodes de bootstrap no paramètric per a obtenir els intervals de confiança.

Però de bootstrap no paramètric n'hi ha més d'un mètode i cabria preguntar-se quin és el més adequat donades les nostres dades. Exposem-los breument,

- El primer mètode es fa servir quan estem fent observacions aleatòries en parelles juntes, és a dir, de la variable resposta i de la variable explicativa, a la vegada. El procediment a seguir per a realitzar aquest mètode és,

1. Estimar el model de regressió lineal entre Y i X.
2. Seleccionem una mostra aleatòria amb reemplaçament de les parelles formades per les variables resposta i explicativa que formen el model de regressió simple, amb la mateixa grandària que la mostra inicial.
3. Ajustem un nou model sobre aquestes dues noves variables simulades.
4. Utilitzem el següent estadístic, que es coneix com l' estadístic t, del mètode bootstrap-t, assumint que l' estimació inicial del paràmetre d' interès és consistent i que la distribució asimptòtica d' aquest estadístic és la Normal estàndard,

$$t_b = \frac{\widehat{\alpha}_b - \hat{\alpha}}{\widehat{sd}_b(\hat{\alpha})},$$

on el subíndex b indica el número de mostra aleatòria en el que obtenim l' estadístic, $\widehat{\alpha}_b$ és l' estimació del pendent en una mostra determinada, $\widehat{sd}_b(\hat{\alpha})$ la seva desviació estàndard i $\hat{\alpha}$ l' estimació del pendent inicial.

5. Repetim els passos del 2 al 4 tantes vegades com considerem necessàries tot estimant, cada vegada, el model de regressió lineal. Finalment obtenim l' interval de confiança del pendent amb un 95% de confiança utilitzant els quantils 0.025 i 0.975 de l' estadístic t.

- El segon mètode, en canvi, es fa servir quan provem diferents valors per a la variable explicativa per a determinar la variable resposta d' interès. Els passos a seguir, en aquest cas, són,

1. Estimar el model de regressió lineal entre Y i X.
2. Seleccionem una mostra aleatòria amb reemplaçament dels errors obtinguts pel model de regressió lineal simple anterior, amb la mateixa grandària que la mostra inicial.
3. Simulem la variable resposta d' interès com,

$$Y = \widehat{\beta}_0 + \widehat{\beta}_1 X + \varepsilon$$

On ε equival a l' estimació dels errors obtinguda en el segon pas.

4. Tornem a estimar el model de regressió lineal, utilitzant la variable Y simulada en el pas anterior.
5. Utilitzem l' estadístic t, esposat al mètode anterior.
6. Repetim els passos del 2 al 5 tantes vegades com considerem necessàries per a obtenir l' interval de confiança del pendent amb un 95% de confiança utilitzem els quantils 0.025 i 0.975 de l' estadístic t.

El primer mètode és menys sensible a assumpcions que el segon, però en el nostre cas, és prou clar que les observacions de totes dues variables no s' observen aleatòriament i a la vegada, si no que nosaltres fixem els valors de la variable explicativa. Com que tots dos mètodes tenen les seves avantatges i desavantatges el millor és calcular els intervals de confiança mitjançant tots dos mètodes i després comparar els resultats¹⁰, que podem observar a la següent taula,

	$\hat{\alpha}$	Primer mètode	Segon mètode
PDI Atlàntic	1.43	(1.68, 1.07)	(1.67, 1.21)
PDI Pacífic	1.108	(1.34, 0.92)	(1.28, 0.92)
IEI Atlàntic	1.23	(1.33, 1.17)	(1.29, 1.17)
IEI Pacífic	1.32	(1.50, 1.22)	(1.42, 1.22)
OEI Atlàntic	1.327	(1.41, 1.26)	(1.39, 1.25)
OEI Pacífic	1.286	(1.45, 1.10)	(1.45, 1.13)

Taula 5: Estimació dels exponents de potència i els seus intervals de confiança.

Sembla important notar que no s' aprecien grans diferències entre els intervals de confiança obtinguts pels dos mètodes. Concretament, com que els intervals de confiança es solapen, no podem dir que hi hagi diferències estadísticament significatives entre les estimacions dels diferents exponents de potència. Això ens indica que, no únicament no trobem diferències rellevants entre ciclons del Atlàntic i del Pacífic, si no que, globalment, l' estimació de l' exponent de potències tampoc mostra diferències significatives entre els tres índexs utilitzats.

4.4 Estimació de la funció de densitat amb el mètode Kernel

El *mètode Kernel* és un mètode no paramètric per a estimar la funció de densitat d' una variable aleatòria de la que desconexim la seva distribució. L' estimació de la funció de densitat a través d' aquest mètode s' obté a través del següent procediment,

Obtenim x_1, x_2, \dots, x_n mostres independents i idènticament distribuïdes d' una distribució amb funció de densitat, f , que estimem com,

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

On K és el que es coneix com a Kernel, una funció simètrica però no necessàriament positiva la integral de la qual és 1, i h fa referència al *bandwidth*¹¹ que no és més que un paràmetre, positiu, suavitzador.

¹⁰ Per a obtenir aquests intervals de confiança s' ha utilitzat un número de repeticions igual a 1000.

¹¹ Per a ampliar l' informació sobre aquest concepte, es pot consultar la pàgina 45 de B.N. Silverman, Density Estimation for Statistics and Data Analysis. Chapman & Hall, 1992.

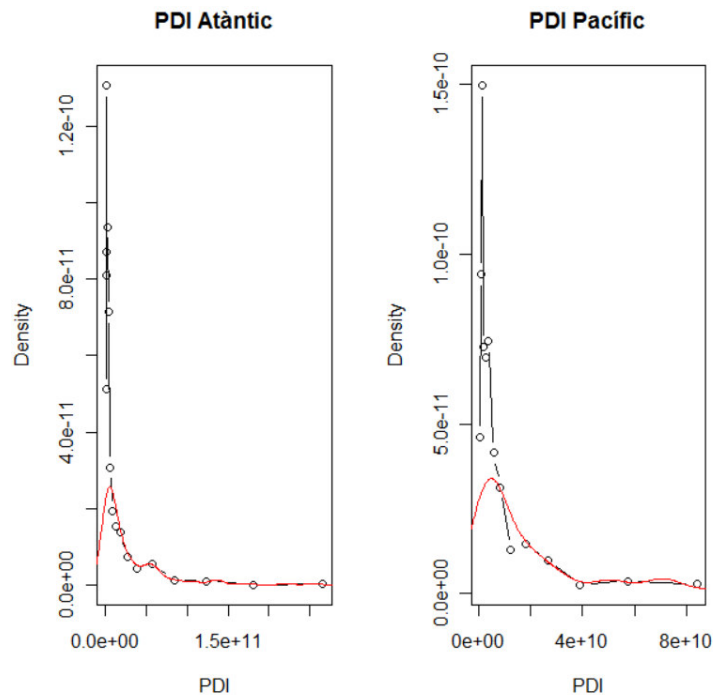
A R trobem una funció implementada capaç d'estimar la funció de densitat d'una determinada variable aleatòria mitjançant el mètode Kernel. Aquesta funció s'anomena *density()* i serà la que utilitzarem per a obtenir la nova estimació de la funció de densitat. Concretament, aquesta funció utilitza el mètode Kernel Normal, que de fet, és el més utilitzat degut a les seves bones propietats matemàtiques. Aquest mètode consisteix a establir el Kernel d'un determinat valor utilitzant una distribució de densitat Normal estàndard (ϕ),

$$K(x) = \phi(x)$$

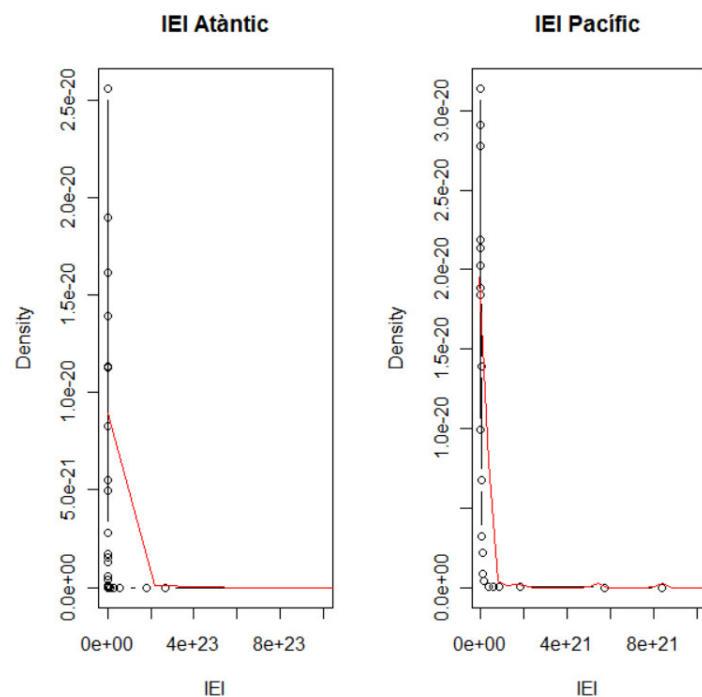
També podem veure que, per defecte, aquesta funció calcula quin *bandwidth* convé utilitzar com a 0.9 vegades el mínim entre la desviació estàndard i el rang interquartílic dividit entre 1.34 vegades la mida mostral. És a dir,

$$0.9(\min\{sd, \frac{IQ}{1.34n}\})$$

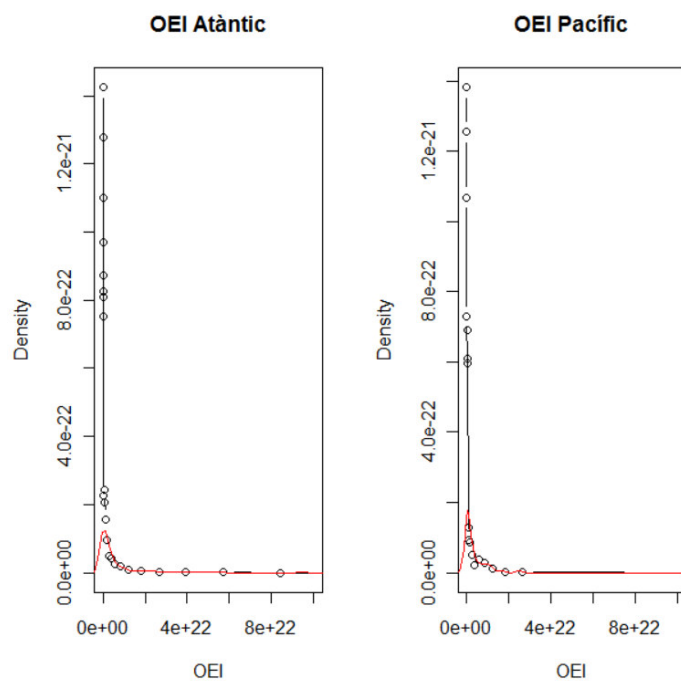
Tenint clar això, el que farem serà tornar a estimar les funcions de densitat dels índexs per tal de comparar els resultats amb els que obteníem a la secció 4.2. A tal efecte, el més pràctic és observar els següents gràfics on trobem la funció de densitat estimada a través del mètode Kernel sobreposada a la funció de densitat estimada al apartat 4.2. En aquest cas, però, utilitzarem la funció de densitat sense aplicar logaritmes, ja que en aquest cas, l'ajust de les dues densitats s'observa millor, i, tal com fèiem per a obtenir la funció de densitat inicial, utilitzem totes les dades, és a dir, també les dades inicials que traïem per a realitzar l'estimació de l'exponent de potències.



Gràfic 16: Comparació de la densitat obtinguda amb la funció Kernel (línia vermella) amb la prèviament calculada per a l'índex PDI.



Gràfic 17: Comparació de la densitat obtinguda amb la funció Kernel (línia vermella) amb la prèviament calculada per a l'índex IEI.



Gràfic 18: Comparació de la densitat obtinguda amb la funció Kernel (línia vermella) amb la prèviament calculada per a l'índex OEI.

En aquests gràfics podem veure que, tal com esperàvem, les dues gràfiques de densitat s'ajusten prou bé un cop passades les primeres dades. Especialment, en el cas dels índexs PDI i OEI.

Per altra banda, hem provat altres models de kernel disponibles a la funció *density()*, és a dir, que utilitzen altres distribucions diferents a la Normal, i també altres valors *bandwidth*, majors i menors al que s'utilitza per defecte. En qualsevol cas, els gràfics que es presenten, amb els paràmetres per defecte, són en els que s'observa un millor ajust de les dues funcions de densitat, en general, per a tots els índexs.

4.5 Test de bondat d'ajust

En aquest últim apartat es vol veure si es pot assumir que els índexs amb els que hem estat treballant segueixen una distribució de Pareto(θ, a), ja que es pot veure que la seva funció de densitat és molt similar a la que hem obtingut aplicant la llei de potències. Concretament, la funció de densitat d'una distribució Pareto(θ, a) es defineix de la següent manera, per a tota x major que el paràmetre a ,

$$\theta a^\theta x^{-(\theta+1)} 1_{[a, +\infty]}$$

I és zero altrament. Mentre que la funció de densitat que segueixen els nostres índexs, sota la llei de potències, es defineix de la següent manera, sempre que x sigui major que un determinat valor A ,

$$C x^{-\alpha} 1_{[A, +\infty]}$$

I també és zero altrament. D'aquesta manera es veu bastant clarament que el paràmetre α amb el que hem estat treballant fins ara és equivalent a $\theta + 1$, on θ és un dels paràmetres que defineix la distribució Pareto, i el paràmetre a fa referència al valor a partir del qual la densitat no és nul·la, és a dir, traient les primeres observacions que desviaven la tendència de les dades, que utilitzant la notació de la llei de potències, anomenem A .

Amb això clar, passem a aplicar el test estadístic de bondat d'ajust. El més habitual és el test de Kolmogorov-Smirnov, però aquest no es pot fer servir directament si els paràmetres que s'utilitzen per a la distribució sota la que pensem que es distribueixen les nostres dades no han estat fixats prèviament, és a dir, aquests paràmetres no poden haver estat estimats de les dades, com és el nostre cas.

Per tant utilitzarem el test d'Anderson-Darling. Aquest test és anàleg al test de Kolmogorov-Smirnov, però amb la particularitat que es pot fer servir en el cas de que els paràmetres de la distribució s'hagin estimat a partir de les dades. Aquest test es troba implementat en R, a la funció *gofGENPARtest()*, que realitza el test d'Anderson-Darling de bondat d'ajust mitjançant el mètode de Monte Carlo.

Concretament, donada una mostra de les dades inicials x_i , amb i entre 1 i m , si aquestes segueixen una determinada distribució, $FD(x)$, aquest test es basa en contrastar l'hipòtesis nul·la de que $FD(x) = F(x, \theta)$, on $F(x, \theta)$ és la hipotètica distribució i els seus paràmetres són estimats a través de la mostra x_i .

El test d'Anderson-Darling mesura la diferència entre la funció de distribució hipotètica i la funció de freqüències acumulades, que seguint l'ajuda de la funció *gofGENPARtest()* d'R anomenarem $F_m(x)$, definida com,

$$\begin{aligned} & 0 \text{ per a } x < x_1, \\ & \frac{i}{m} \text{ per a } x_i \leq x < x_{i+1}, \\ & 1 \text{ per a } x_m \leq x, \end{aligned}$$

on x_i és l'element i -èssim de la mostra ordenada en sentit creixent.

L'estadístic que utilitza aquest test és el següent,

$$A^2 = -m - \frac{1}{m} \sum_{i=1}^m (2i-1) / \ln F(x_i, \theta) + (2m+1-2i) / \ln (1-F(x_i, \theta))$$

Per tant, apliquem aquesta funció a les nostres dades per a obtenir els p-valors del test, que podem observar a la taula 6. Cal notar, però, que és adient, com hem comentat prèviament, treure les dades inicials, que considerem que no segueixen la llei de potències. Per a trobar el valor a partir del qual es pot considerar que les dades segueixen la llei de potències, utilitzem dos procediments. El primer, consisteix a agafar el valor a partir del qual considerem que les dades tenen el comportament donat per la llei de potències i que s'havia utilitzat a l'estimació de l'exponent per mitja de l'estimació del pendent de les rectes de regressió del logaritme de les funcions de densitat, a aquest valor estimat, l'anomenarem \widehat{a}_1 . El segon mètode, consisteix a estimar aquests paràmetre gràcies als valors $\widehat{\beta}_0$ i $\widehat{\beta}_1$ de les mateixes regressions. Per a fer-ho, utilitzem l'assumpció de que l'integral de la funció de densitat de la llei de potències ha de ser igual a 1, de manera que trobem la següent igualtat,

$$C = (\alpha - 1)a^{\alpha-1}$$

A més sabem que, al realitzar la regressió entre la variable resposta d'interès, el logaritme de la densitat estimada, i la variable explicativa, el logaritme dels punts centrals entre dos determinats breaks, es donen les següents igualtats,

$$C = e^{\widehat{\beta}_0}$$

$$\alpha = \widehat{\beta}_1$$

Amb tot això clar, podem establir que,

$$a_2 = \frac{e^{\widehat{\beta}_0}}{\sqrt{-\widehat{\beta}_1 - 1}}^{-\widehat{\beta}_1 - 1}$$

Per tant, apliquem el test de bondat d' ajust d' Anderson-Darling, utilitzant el paràmetre a estimat pels dos mètodes. Podem trobar els p-valors obtinguts en cada cas a la següent taula,

	\widehat{a}_1	Primer mètode	\widehat{a}_2	Segon mètode
PDI Atlàntic	$4.641589 \cdot 10^8$	0.0895	$1.747279327 \cdot 10^9$	0.1305
PDI Pacífic	$6.812921 \cdot 10^8$	0.1325	$1.055118972 \cdot 10^{13}$	NA
IEI Atlàntic	$4.641589 \cdot 10^{17}$	0	$3.030056419 \cdot 10^{18}$	0.001
IEI Pacífic	$1.467799 \cdot 10^{18}$	0.006	$4.862977242 \cdot 10^{18}$	0.0045
OEI Atlàntic	$1 \cdot 10^{20}$	0.025	$8.954468458 \cdot 10^{19}$	0.03
OEI Pacífic	$1.467799 \cdot 10^{20}$	0.118	$2.266191063 \cdot 10^{20}$	0.0345

Taula 6: P-valors del test d' Anderson-Darling de bondat d' ajust per a una distribució hipotètica Pareto (θ, a) .

Aquesta taula ens mostra que, agafant un nivell de confiança de l' 1%, podem pensar que els índexs si que segueixen una distribució de Pareto, ja sigui per al primer mètode, per al segon, o per a tots dos, en tots els casos llevat del cas de l' índex IEI. Notem també, que un dels p-valors apareix com a dada faltant, això és degut a que l' estimació del paràmetre a de la distribució de Pareto en aquest cas dóna un valor massa gran, i no hi ha cap valor d' aquest índex major a aquest valor, per tant, aquest p-valor no és calculable.

L' únic inconvenient que presenta aquest test és que estima els paràmetres a traves de les dades, és a dir, no podem especificar que utilitzi les estimacions que nosaltres hem obtingut. Això, no obstant, no té perquè suposar una gran limitació. Es coneix que el mètode més habitual per a estimar el paràmetre θ d' una distribució de Pareto és a traves del mètode de la màxima versemblança, concretament, s' utilitza la següent estimació,

$$\hat{\theta}_{MV} = \frac{n}{\sum_1^n \log \frac{x_i}{a}},$$

on n és el número d' observacions que s' observem. Utilitzant aquesta estimació, podem obtenir les estimacions del paràmetre θ de la distribució de Pareto, com abans, utilitzant dos mètodes per a obtenir el paràmetre a , i així, obtenim l' estimació del paràmetre a de la llei de potències. Podem observar els resultats a la següent taula, on el subíndex 1, indica que hem utilitzat el paràmetre a obtingut pel primer mètode, i el dos indica que hem utilitzat l' obtingut mitjançant el segon mètode,

	$\hat{\theta}_{MV1}$	$\hat{\alpha}_{MV1}$	$\hat{\theta}_{MV2}$	$\hat{\alpha}_{MV2}$
PDI Atlàntic	0.319	1.319	0.462	1.462
PDI Pacífic	0.377	1.377	NA	NA
IEI Atlàntic	0.193	1.193	0.295	1.295
IEI Pacífic	0.309	1.309	0.453	1.453
OEI Atlàntic	0.331	1.331	0.321	1.321
OEI Pacífic	0.395	1.395	0.442	1.442

Taula 7: Estimacions del paràmetre θ de la distribució de Pareto, i del paràmetre α de la llei de potències amb el mètode de la màxima versemblança.

Gràcies a aquesta taula podem veure que les noves estimacions dels paràmetres de la llei de potències es troben dins dels intervals de confiança de les estimacions obtingudes prèviament, o en qualsevol cas no molt allunyats. Això ens indica que si realitzéssim intervals de confiança per a aquest nou mètode d'estimació, aquests es solaparien amb els intervals de confiança calculats prèviament, i per tant, en qualsevol cas, no podríem dir que existeixen diferències estadísticament significatives entre aquests dos mètodes d'estimació per al nostre cas.

5. Conclusions

Gràcies a l'anàlisi inicial de les sèries del número de ciclons anuals i mensuals de l'Atlàntic i el Pacífic hem pogut veure que no s'observa una relació entre el número de ciclons que es donen entre un any i el següent, o l'anterior, però, el que sí que s'observa és una diferenciació entre la sèrie del número de ciclons anuals de l'Atlàntic i el Pacífic, ja que el número de ciclons anuals a l'Atlàntic presenta un petit pendent creixent, el que ens indica que el número de ciclons va augmentant entre un any i el següent, i això, en canvi, no s'observa pel que fa al Pacífic.

Per altra banda, després de l'anàlisi de les sèries mensuals del número de ciclons, concloem que un model apropiat per a la sèrie del número de ciclons de l'Atlàntic segueix un procés SARIMA(1,1,1) x (0,1,1)₁₂, mentre que per a la del Pacífic seria un procés SARIMA(0,1,1) x (1,1,1)₁₂.

Un altre resultat que obtenim gràcies a l'anàlisi de les sèries del número de ciclons mensuals, és l'influència de la temperatura, concretament, s'observa un augment del número de ciclons mensuals com a conseqüència d'un augment de la temperatura en un instant del temps anterior, en aquest cas, doncs, parlem de mesos, per tant, després de que un determinat mes presenti un augment en la temperatura, esperarem que el número de ciclons que s'observen al mes següent augmenti. Cal aclarir que això és un resultat únicament visual, de fet, la relació de les sèries dels números de ciclons amb les temperatures és un problema interessant però que no s'ha abordat a fons en aquest treball.

Finalment, a través de l'anàlisi dels índexs meteorològics PDI, IEI i OEI obtenim diversos resultats a mencionar. El primer és el fet de que el logaritme d'aquests índexs es trobin molt correlats. A més, a l'aplicar un histograma amb dues escales logarítmiques obtenim rectes de les quals estimem el pendent per mitjà d'una regressió. Aquest comportament és típic de les lleis de potències fent el seu exponent el pendent de la recta considerada. Hem també utilitzat el mètode Kernel gaussià per a estimar la densitat obtenint resultats molt semblants. Per altra banda, podem concloure, en pràcticament tots els casos, que els índexs provenen d'una distribució de Pareto, amb paràmetres estimats directament de les nostres dades. Per últim, un dels resultats més interessants d'aquest tema és el fet de que no trobem diferències estadísticament significatives entre les estimacions dels paràmetres de la llei de potències estimats a partir dels diferents índexs, ni tampoc entre els índexs de l'Atlàntic i el Pacífic. Aquest resultat, unit al fet de la elevada correlació observada entre el logaritme dels índexs, ens fa pensar que ***tot i que són tres índexs diferents, i es construeixen de diferent manera, com que es fan servir per al mateix, és molt possible que únicament sigui necessari, a la pràctica, fer servir un d'ells, ja que segons els resultats obtinguts, no s'esperaria obtenir moltes diferències estadísticament rellevants entre ells.***

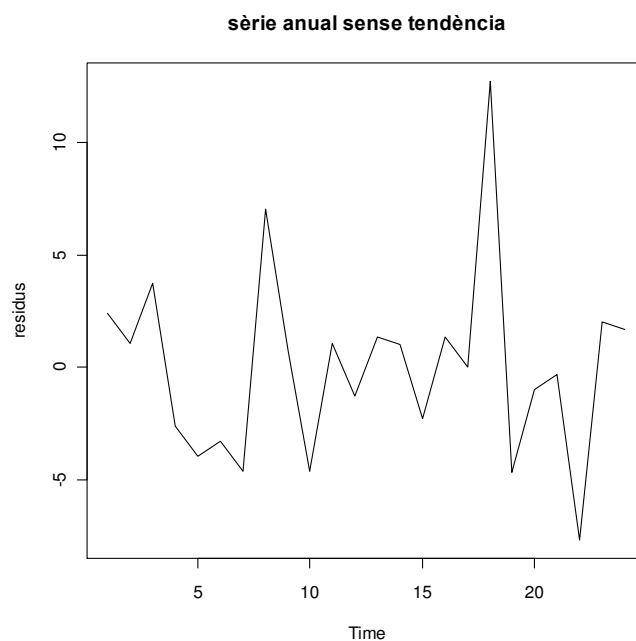
ANNEXOS

Annex A1:

Lag	Atlàntic	Pacífic
1	0.6628	0.4094
2	0.7592	0.4047
3	0.6258	0.2313
4	0.7545	0.2129
5	0.3995	0.2979
6	0.4598	0.3884
7	0.5766	0.4372
8	0.6395	0.5008
9	0.7297	0.594
10	0.7194	0.6734
11	0.495	NA
12	0.5104	NA
13	0.412	NA
14	0.1949	NA
15	0.2232	NA
16	0.2516	NA
17	0.242	NA
18	0.2671	NA
19	0.2039	NA
20	0.2417	NA
21	0.2917	NA
22	0.3198	NA
23	0.3672	NA

Taula A1: P-valors obtinguts per al Test Box Ljung per a les sèries anuals del Atlàntic i Pacífic.

Annex A2:

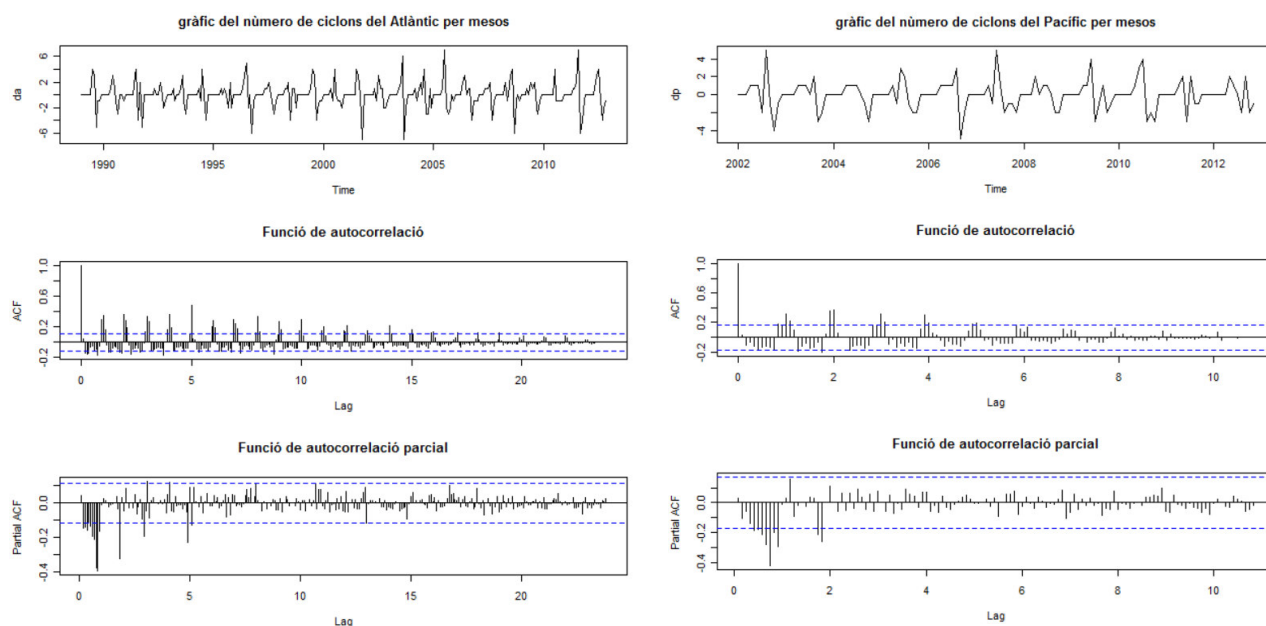


Gràfic A1: Sèrie de número de ciclons anual del Atlàntic sense tendència.

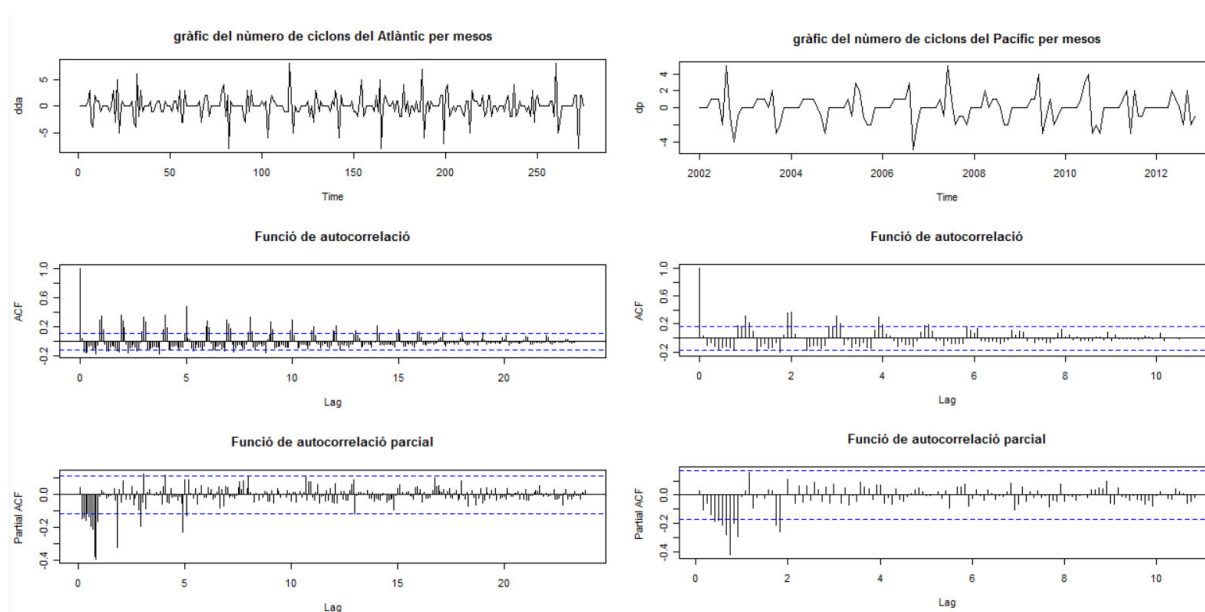
Lag	Atlàntic
1	0.3655
2	0.5965
3	0.7762
4	0.2919
5	0.2658
6	0.3746
7	0.4818
8	0.5069
9	0.6078
10	0.3836
11	0.2725
12	0.3457
13	0.382
14	0.3163
15	0.1143
16	0.1394
17	0.1524
18	0.1931
19	0.1842
20	0.2274
21	0.2708
22	0.3189
23	0.3704

Taula A2: P-valors obtinguts per al Test Box Ljung per a les sèries anuals del Atlàntic sense tendència.

Annex A3:



Gràfic A3: Sèrie de número de ciclons mensual del Atlàntic i del Pacífic després d' aplicar l' operador diferència una vegada.



Gràfic A4: Sèrie de número de ciclons mensual del Atlàntic i del Pacífic sense estacionalitat.

Annex A4:

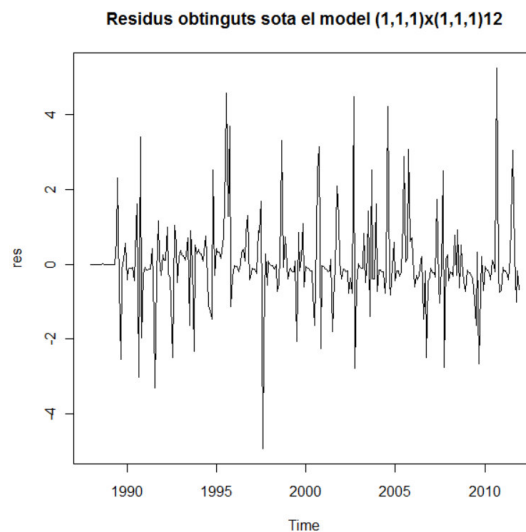
Procés	Atlàntic	Pacífic
$(1,1,1) \times (1,1,1)_{12}$	927.87	377.34
$(0,1,1) \times (1,1,1)_{12}$	952.13	380.44
$(1,1,0) \times (1,1,1)_{12}$	1000.14	416.76
$(0,1,0) \times (1,1,1)_{12}$	1031.09	426.23
$(1,1,1) \times (0,1,1)_{12}$	927.76	378.43
$(1,1,1) \times (1,1,0)_{12}$	1024.35	386.53
$(1,1,1) \times (0,1,0)_{12}$	1095.78	425.93
$(0,1,1) \times (0,1,1)_{12}$	954.24	381.83
$(0,1,1) \times (0,1,0)_{12}$	1121.42	431.8
$(0,1,1) \times (1,1,0)_{12}$	1045.81	391.28
$(1,1,0) \times (0,1,1)_{12}$	1000.79	418.93
$(1,1,0) \times (0,1,0)_{12}$	1174.98	468.76
$(1,1,0) \times (1,1,0)_{12}$	1096.87	426.65
$(0,1,0) \times (0,1,1)_{12}$	1030.82	427.33
$(0,1,0) \times (1,1,0)_{12}$	1138.56	435.16

Taula A3: AIC obtinguts per als diferents models SARIMA per a les sèries mensuals del Atlàntic i Pacífic.

Annex A5:

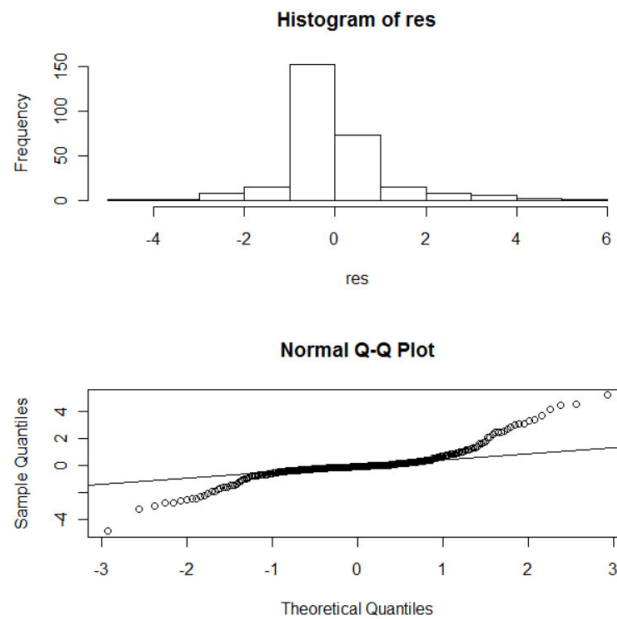
1. Resum gràfic de l' anàlisi dels residus obtinguts aplicant el model SARIMA(1,1,1) x (1,1,1)₁₂ per a la sèrie del Atlàntic.

El primer gràfic mostra els residus obtinguts, observem que es distribueixen caòticament al voltant del zero. Això es una bona senyal, indica que els errors són incorrelats i possiblement amb mitjana igual a zero. Observem, també, poca variabilitat.



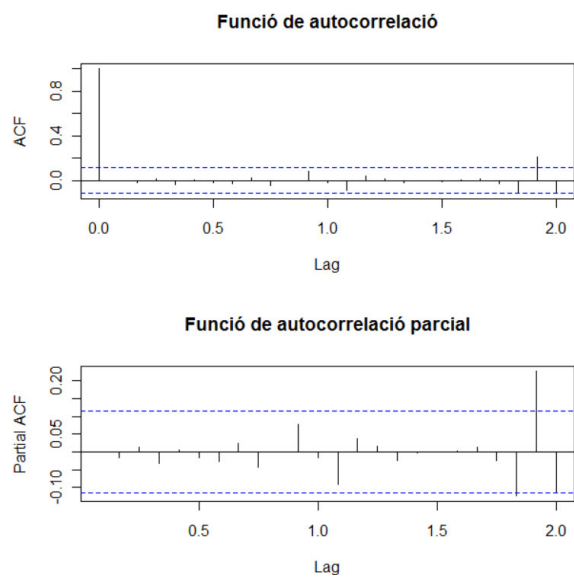
Gràfic A5: Residus obtinguts aplicant el model SARIMA(1,1,1) x (1,1,1)₁₂ per a la sèrie del Atlàntic.

Per altra banda, podem visualitzar un histograma i un qqplot per a contrastar l' hipòtesis de normalitat visualment. D' aquesta manera veiem que els residus presenten un problema de cues, és dir, les probabilitats dels valors allunyats del centre no corresponen a les que haurien de ser sota la hipòtesis de normalitat.



Gràfic A6: Histograma i un qqplot dels residus obtinguts aplicant el model SARIMA(1,1,1) x (1,1,1)₁₂ per a la sèrie del Atlàntic.

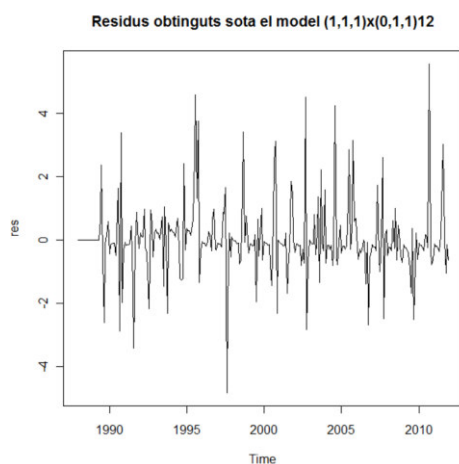
Finalment, visualitzem el gràfic d' autocorrelació parcial, on veiem que els residus si tenen la forma d' un soroll blanc.



Gràfic A7: Funció d' autocorrelació i d' autocorrelació parcial dels residus obtinguts aplicant el model SARIMA(1,1,1) x (1,1,1)₁₂ per a la sèrie del Atlàntic.

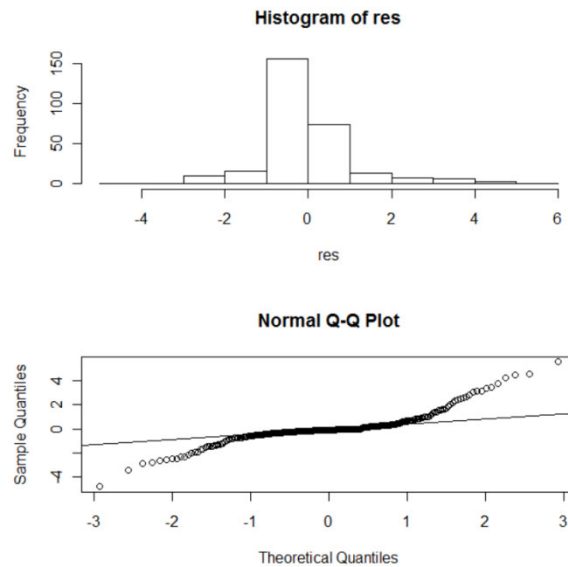
2. Resum gràfic de l' anàlisi dels residus obtinguts aplicant el model SARIMA(1,1,1) x (0,1,1)₁₂ per a la sèrie del Atlàntic.

Seguint el mateix procediment, el primer gràfic mostra els residus obtinguts, observem que també podem pensar que els errors són incorrelats i possiblement amb mitjana igual a zero i amb poca variabilitat.



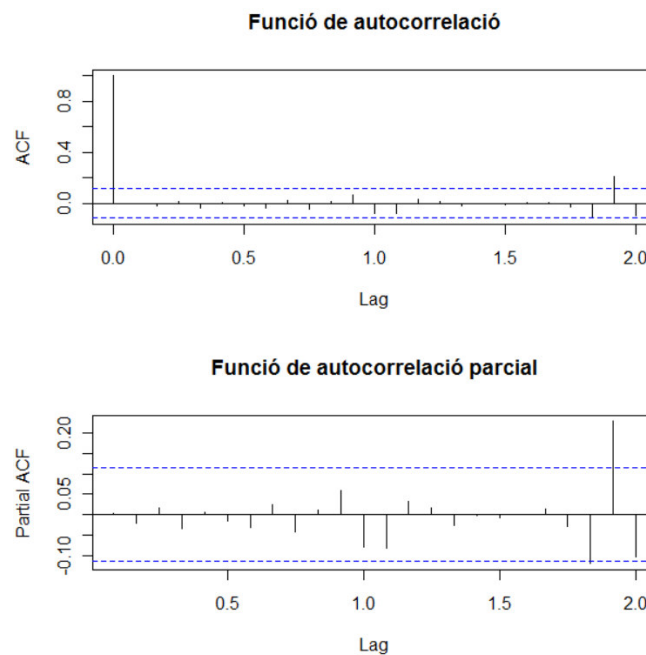
Gràfic A8: Residus obtinguts aplicant el model SARIMA(1,1,1) x (0,1,1)₁₂ per a la sèrie del Atlàntic.

Al visualitzar l' histograma i el qqplot, obtenim unes conclusions molt similars a les de abans.



Gràfic A9: Histograma i qqplot dels residus obtinguts aplicant el model $\text{SARIMA}(1,1,1) \times (1,1,1)_{12}$ per a la sèrie del Atlàntic.

Finalment, visualitzem el gràfic d' autocorrelació parcial, on veiem que els residus, en aquest cas, també tenen la forma d' un soroll blanc.

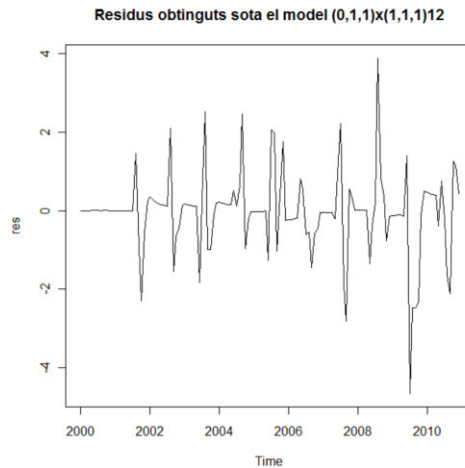


Gràfic A10: Funció d' autocorrelació i d' autocorrelació parcial dels residus obtinguts aplicant el model $\text{SARIMA}(1,1,1) \times (1,1,1)_{12}$ per a la sèrie del Atlàntic.

Annex A6:

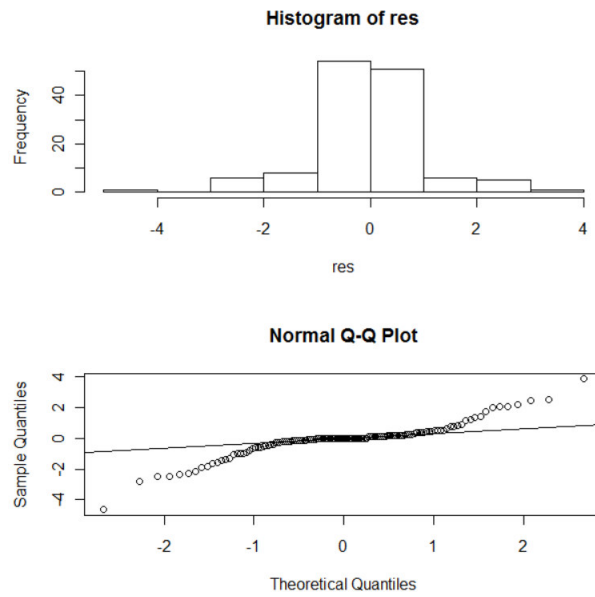
Resum gràfic de l'anàlisi dels residus obtinguts aplicant el model SARIMA(0,1,1) x (1,1,1)₁₂ per a la sèrie del Pacífic.

El primer gràfic mostra els residus obtinguts, en aquest cas també podem dir que els errors semblen incorrelats, possiblement amb mitjana igual a zero i amb poca variabilitat.



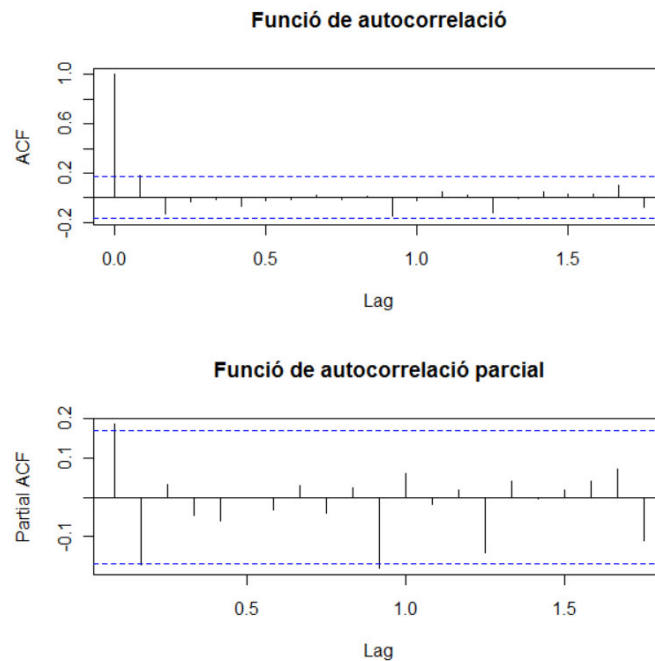
Gràfic A11: Residus obtinguts aplicant el model SARIMA(0,1,1) x (1,1,1)₁₂ per a la sèrie del Pacífic.

L'histograma i el qqplot ens mostren que tornem a tenir, en aquest cas, problemes de cues.



Gràfic A12: Histograma i qqplot dels residus obtinguts aplicant el model $SARIMA(0,1,1) \times (1,1,1)_{12}$ per a la sèrie del Pacífic.

Finalment, visualitzem el gràfic d' autocorrelació parcial, on veiem que els residus tornen a tenir la forma d' un soroll blanc.



Gràfic A13: Funció d' autocorrelació i d' autocorrelació parcial dels residus obtinguts aplicant el model $SARIMA(0,1,1) \times (1,1,1)_{12}$ per a la sèrie del Pacífic.

Annex A7: Sintaxis d' R

#Introduim les dades

```
atlc <- sqlQuery(channel = 5, select * from [Hoja1$])
names(atlc) <- make.names(names(atlc))
```

```
pac <- sqlQuery(channel = 6, select * from [Hoja1$])
names(pac) <- make.names(names(pac))
```

#Unim les dues bases de dades en una sola

```
localització <- c(rep(1, nrow(atlc)), rep(2, nrow(pac)))
d <- rbind(atlc, pac)
d <- data.frame(d, localització)
```

#Variables categòriques

#Variable tipus

```
d$tipus <- factor(d$tipus)
levels(d$tipus) <- list("sistema tropical" = " * ", "ona tropical" = "W", "perturbació tropical"
                        = "D", "tormenta subtropical" = "S", "tormenta extra - tropical"
                        = "E", "residus baixos" = "L")
```

#Variable localització

```
d$localització <- factor(d$localització)
levels(d$localització) <- list("atlàntic" = 1, "pacífic" = 2)
```

#Dades faltants

```
d[d == -99] <- NA
dadesfaltants <- NULL
pmv <- -100 * apply(is.na(d), 2, mean)
pmi <- -100 * apply(is.na(d), 1, mean)
c <- subset(pmi, pmi == 0)
pcc <- -100 * (length(c)/length(d[,1]))
dadesfaltants$percentatge.de.casos.complets <- pcc
dadesfaltants$percentatge.de.missings.per.variables <- -pmv
dadesfaltants$resum.de.missings.per.variables <- summary(pmv)
dadesfaltants$resum.de.missings.per.individus <- summary(pmi)
dadesfaltants
d <- d[, (pmv < 30)]
```

#Eliminem les variables que no utilitzarem

```
d <- d[, c(1:6, 9, 11, 13, 26, 28)]
```

#Contruïm una sèrie amb el número de ciclons del Atlàntic anuals i mensuals

```
anys <- c(1988:2011)
mesos <- c(1:12)
ciclons.any.atlc <- 0
ciclons.mesos.atlc <- rep(0, 12 * 24)
dim(ciclons.mesos.atlc) <- c(24, 12)
for(i in 1:length(anys))
{
  a <- subset(atlc, any == anys[i])
  b <- summary(a$nom)
```

```

ciclons.any.atlc[i] <- sum(as.numeric(b > 0))

for(j in 1:length(mesos))
{
  m <- subset(a, mes == mesos[j])
  m2 <- summary(m$nom)
  ciclons.mesos.atlc[i,j] <- sum(as.numeric(m2 > 0))
}
}

#Contruïm una sèrie amb el número de ciclons del Pacífic anual i mensual
anys <- c(2001:2011)
mesos <- c(1:12)
ciclons.any.pac <- 0
ciclons.mesos.pac <- rep(0,11 * 12)
dim(ciclons.mesos.pac) <- c(11,12)
for(i in 1:length(anys))
{
  a <- subset(pac, any == anys[i])
  b <- summary(a$nom)
  ciclons.any.pac[i] <- sum(as.numeric(b > 0))

  for(j in 1:length(mesos))
  {
    m <- subset(a, mes == mesos[j])
    m2 <- summary(m$nom)
    ciclons.mesos.pac[i,j] <- sum(as.numeric(m2 > 0))
  }
}

par(mfrow = c(2,2))
#El convertim en una sèrie temporal i indiquem que comenci a l'any 1988
(ciclons.any.atlc = ts(ciclons.any.atlc, start = 1988))
plot.ts(ciclons.any.atlc, main = "gràfic del número de ciclons del atlàntic per any")
ciclons.mesos.atlc <- as.vector(t(ciclons.mesos.atlc))
#Indiquem que comenci a l'any 1988 i va per mesos (12 en un any)
(ciclons.mesos.atlc = ts(ciclons.mesos.atlc, start = c(1988,1), frequency = 12))
plot.ts(ciclons.mesos.atlc, main = "gràfic del número de ciclons del atlàntic per mesos")

#El convertim en una sèrie temporal i indiquem que comenci a l'any 1988
(ciclons.any.pac = ts(ciclons.any.pac, start = 2001))
plot.ts(ciclons.any.pac, main = "gràfic del número de ciclons del pacífic per any")
ciclons.mesos.pac <- as.vector(t(ciclons.mesos.pac))
#Indiquem que comenci a l'any 1988 i va per mesos (12 en un any)
(ciclons.mesos.pac = ts(ciclons.mesos.pac, start = c(2001,1), frequency = 12))
plot.ts(ciclons.mesos.pac, main = "gràfic del número de ciclons del pacífic per mesos")

```

```

par(mfrow = c(3,1))
plot.ts(ciclons.any.atlc, main = "gràfic del número de ciclons del atlàntic per any")
#estimacions de les funcions ACF i PACF de la sèrie
f.autocorm <- acf(ciclons.any.atlc, type = "correlation", plot = FALSE, lag.max
                 = (length(ciclons.any.atlc) - 1)); f.autocorm
f.autocor.parm <- pacf(ciclons.any.atlc, type = "correlation", plot = FALSE, lag.max
                     = (length(ciclons.any.atlc) - 1)); f.autocor.parm
plot(f.autocorm, main = "Funció de autocorrelació")
plot(f.autocor.parm, main = "Funció de autocorrelació parcial")

par(mfrow = c(3,1))
plot.ts(ciclons.mesos.atlc, main = "gràfic del número de ciclons del atlàntic per mesos")
f.autocorm <- acf(ciclons.mesos.atlc, type = "correlation", plot = FALSE, lag.max
                 = (length(ciclons.mesos.atlc) - 1)); f.autocorm
f.autocor.parm <- pacf(ciclons.mesos.atlc, type = "correlation", plot = FALSE, lag.max
                     = (length(ciclons.mesos.atlc) - 1)); f.autocor.parm
plot(f.autocorm, main = "Funció de autocorrelació")
plot(f.autocor.parm, main = "Funció de autocorrelació parcial")

par(mfrow = c(3,1))
(ciclons.any.pac = ts(ciclons.any.pac, start = 2001))
plot.ts(ciclons.any.pac, main = "gràfic del número de ciclons del pacífic per any")
f.autocorm <- acf(ciclons.any.pac, type = "correlation", plot = FALSE, lag.max
                 = (length(ciclons.any.pac) - 1)); f.autocorm
f.autocor.parm <- pacf(ciclons.any.pac, type = "correlation", plot = FALSE, lag.max
                     = (length(ciclons.any.pac) - 1)); f.autocor.parm
plot(f.autocorm, main = "Funció de autocorrelació")
plot(f.autocor.parm, main = "Funció de autocorrelació parcial")

par(mfrow = c(3,1))
plot.ts(ciclons.mesos.pac, main = "gràfic del número de ciclons del pacífic per mesos")
f.autocorm <- acf(ciclons.mesos.pac, type = "correlation", plot = FALSE, lag.max
                 = (length(ciclons.mesos.pac) - 1)); f.autocorm
f.autocor.parm <- pacf(ciclons.mesos.pac, type = "correlation", plot = FALSE, lag.max
                     = (length(ciclons.mesos.pac) - 1)); f.autocor.parm
plot(f.autocorm, main = "Funció de autocorrelació")
plot(f.autocor.parm, main = "Funció de autocorrelació parcial")

#Es poden considerar las sèries dels anys soroll blanc?
library(stats)
Box.test(ciclons.any.atlc, lag = 1, type = c("Ljung"))
Box.test(ciclons.any.pac, lag = 1, type = c("Ljung"))

#Observem tendència lineal?
tlaa <- -data.frame(ciclons.any.atlc, c(1988:2011))
names(tlaa) <- -c("num", "any")
mtlaa <- -lm(num~any, data = tlaa)
summary(mtlaa)

```

```
#Gràfic d' ajust de la tendència lineal
plot.ts(ciclons.any.atlc,main = "gràfic del número de ciclons del Atlàntic per any")
lines(tlaa$any,mtlaa$fitted.values, col = "red")
```

```
#Treiem la tendència lineal per la sèrie del Atlàntic
plot.ts(mtlaa$residuals,main = "sèrie anual sense tendència",ylab = "residus")
Box.test(mtlaa$residuals,lag = 1,type = c("Ljung"))
tlap <- data.frame(ciclons.any.pac,c(2001:2011))
names(tlap) <- c("num","any")
mtlap <- lm(num~any,data = tlap)
summary(mtlap)
```

```
#Apliquem l'operador diferència un cop a les sèries mensuals
par(mfrow = c(3,1))
n <- -length(ciclons.mesos.atlc)
da <- -(ciclons.mesos.atlc[2:n] - ciclons.mesos.atlc[1:(n - 1)])
da = ts(da,start = c(1989,1),frequency = 12)
plot.ts(da,main = "gràfic del número de ciclons del atlàntic per mesos")
f.autocorm <- acf(da,type = "correlation",plot = FALSE,lag.max
= (length(ciclons.mesos.atlc) - 1)); f.autocorm
f.autocor.parm <- pacf(da,type = "correlation",plot = FALSE,lag.max
= (length(ciclons.mesos.atlc) - 1)); f.autocor.parm
plot(f.autocorm,main = "Funció de autocorrelació")
plot(f.autocor.parm,main = "Funció de autocorrelació parcial")
```

```
par(mfrow = c(3,1))
n <- -length(ciclons.mesos.pac)
dp <- -(ciclons.mesos.pac[2:n] - ciclons.mesos.pac[1:(n - 1)])
dp = ts(dp,start = c(2002,1),frequency = 12)
plot.ts(dp,main = "gràfic del número de ciclons del pacífic per mesos")
f.autocorm <- acf(dp,type = "correlation",plot = FALSE,lag.max
= (length(ciclons.mesos.pac) - 1)); f.autocorm
f.autocor.parm <- pacf(dp,type = "correlation",plot = FALSE,lag.max
= (length(ciclons.mesos.pac) - 1)); f.autocor.parm
plot(f.autocorm,main = "Funció de autocorrelació")
plot(f.autocor.parm,main = "Funció de autocorrelació parcial")
```

```
#Ciclons mesos sense estacionalitat
par(mfrow = c(3,1))
n <- -length(ciclons.mesos.atlc)
dda <- -(da[13:n] - da[1:(n - 12)])
plot.ts(dda,main = "gràfic del número de ciclons del Atlàntic per mesos")
f.autocorm <- acf(da,type = "correlation",plot = FALSE,lag.max
= (length(ciclons.mesos.atlc) - 1)); f.autocorm
f.autocor.parm <- pacf(da,type = "correlation",plot = FALSE,lag.max
= (length(ciclons.mesos.atlc) - 1)); f.autocor.parm
plot(f.autocorm,main = "Funció de autocorrelació")
plot(f.autocor.parm,main = "Funció de autocorrelació parcial")
```



```

par(mfrow = c(3,1))
n <- -length(ciclons.mesos.pac)
ddp <- -(dp[13:n] - dp[1:(n - 12)])
plot.ts(dp, main = "gràfic del nombre de ciclons del Pacífic per mesos")
f.autocorm <- acf(dp, type = "correlation", plot = FALSE, lag.max
= (length(ciclons.mesos.pac) - 1)); f.autocorm
f.autocor.parm <- pacf(dp, type = "correlation", plot = FALSE, lag.max
= (length(ciclons.mesos.pac) - 1)); f.autocor.parm
plot(f.autocorm, main = "Funció de autocorrelació")
plot(f.autocor.parm, main = "Funció de autocorrelació parcial")

```

#Estudi de les sèries mensuals

#Atlantic

```

fit <- -list()
fit[[1]] <- -arima(ciclons.mesos.atlc, order = c(1,1,1), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[2]] <- -arima(ciclons.mesos.atlc, order = c(0,1,1), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[3]] <- -arima(ciclons.mesos.atlc, order = c(1,1,0), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[4]] <- -arima(ciclons.mesos.atlc, order = c(0,1,0), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[5]] <- -arima(ciclons.mesos.atlc, order = c(1,1,1), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[6]] <- -arima(ciclons.mesos.atlc, order = c(1,1,1), seasonal = list(order
= c(1, 1, 0), period = 12))
fit[[7]] <- -arima(ciclons.mesos.atlc, order = c(1,1,1), seasonal = list(order
= c(0, 1, 0), period = 12))
fit[[8]] <- -arima(ciclons.mesos.atlc, order = c(0,1,1), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[9]] <- -arima(ciclons.mesos.atlc, order = c(0,1,1), seasonal = list(order
= c(0, 1, 0), period = 12))
fit[[10]] <- -arima(ciclons.mesos.atlc, order = c(0,1,1), seasonal = list(order
= c(1, 1, 0), period = 12))
fit[[11]] <- -arima(ciclons.mesos.atlc, order = c(1,1,0), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[12]] <- -arima(ciclons.mesos.atlc, order = c(1,1,0), seasonal = list(order
= c(0, 1, 0), period = 12))
fit[[13]] <- -arima(ciclons.mesos.atlc, order = c(1,1,0), seasonal = list(order
= c(1, 1, 0), period = 12))
fit[[14]] <- -arima(ciclons.mesos.atlc, order = c(0,1,0), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[15]] <- -arima(ciclons.mesos.atlc, order = c(0,1,0), seasonal = list(order
= c(1, 1, 0), period = 12))

```

#Anàlisi dels residus

#Residus

```

res <- -residuals(fit[[1]])
plot(res, main = "Residus obtinguts sota el model (1,1,1)x(1,1,1)12")
par(mfrow = c(2,1))
hist(res)
qqnorm(res); qqline(res)

```

```

shapiro.test(res)
f.autocor.res <- acf(res, type = "correlation"); f.autocor.res
f.autocor.par.res <- pacf(res, type = "correlation"); f.autocor.par.res
par(mfrow = c(2,1))
plot(f.autocor.res, main = "Funció de autocorrelació")
plot(f.autocor.par.res, main = "Funció de autocorrelació parcial")
library(stats)
Box.test(res, lag = 4, type = c("Ljung"))
tsdiag(fit[[1]], gof.lag = 20)

```

```

res <- residuals(fit[[5]])
plot(res, main = "Residus obtinguts sota el model (1,1,1)x(0,1,1)12")
par(mfrow = c(2,1))
hist(res)
qqnorm(res); qqline(res)
shapiro.test(res)
f.autocor.res <- acf(res, type = "correlation"); f.autocor.res
f.autocor.par.res <- pacf(res, type = "correlation"); f.autocor.par.res
par(mfrow = c(2,1))
plot(f.autocor.res, main = "Funció de autocorrelació")
plot(f.autocor.par.res, main = "Funció de autocorrelació parcial")
library(stats)
Box.test(res, lag = 4, type = c("Ljung"))
tsdiag(fit[[5]], gof.lag = 20)

```

```

#Pacífic
fit <- list()
fit[[1]] <- arima(ciclons.mesos.pac, order = c(1,1,1), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[2]] <- arima(ciclons.mesos.pac, order = c(0,1,1), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[3]] <- arima(ciclons.mesos.pac, order = c(1,1,0), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[4]] <- arima(ciclons.mesos.pac, order = c(0,1,0), seasonal = list(order
= c(1, 1, 1), period = 12))
fit[[5]] <- arima(ciclons.mesos.pac, order = c(1,1,1), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[6]] <- arima(ciclons.mesos.pac, order = c(1,1,1), seasonal = list(order
= c(1, 1, 0), period = 12))
fit[[7]] <- arima(ciclons.mesos.pac, order = c(1,1,1), seasonal = list(order
= c(0, 1, 0), period = 12))
fit[[8]] <- arima(ciclons.mesos.pac, order = c(0,1,1), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[9]] <- arima(ciclons.mesos.pac, order = c(0,1,1), seasonal = list(order
= c(0, 1, 0), period = 12))
fit[[10]] <- arima(ciclons.mesos.pac, order = c(0,1,1), seasonal = list(order
= c(1, 1, 0), period = 12))
fit[[11]] <- arima(ciclons.mesos.pac, order = c(1,1,0), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[12]] <- arima(ciclons.mesos.pac, order = c(1,1,0), seasonal = list(order
= c(0, 1, 0), period = 12))

```

```

fit[[13]] <- arima(ciclons.mesos.pac, order = c(1,1,0), seasonal = list(order
= c(1, 1, 0), period = 12))
fit[[14]] <- arima(ciclons.mesos.pac, order = c(0,1,0), seasonal = list(order
= c(0, 1, 1), period = 12))
fit[[15]] <- arima(ciclons.mesos.pac, order = c(0,1,0), seasonal = list(order
= c(1, 1, 0), period = 12))

```

```

res <- residuals(fit[[2]])
plot(res, main = "Residus obtinguts sota el model (0,1,1)x(1,1,1)12")
par(mfrow = c(2,1))
hist(res)
qqnorm(res); qqline(res)
shapiro.test(res)
f.autocor.res <- acf(res, type = "correlation"); f.autocor.res
f.autocor.par.res <- pacf(res, type = "correlation"); f.autocor.par.res
par(mfrow = c(2,1))
plot(f.autocor.res, main = "Funció de autocorrelació")
plot(f.autocor.par.res, main = "Funció de autocorrelació parcial")
library(stats)
Box.test(res, lag = 4, type = c("Ljung"))
tsdiag(fit[[1]], gof.lag = 20)

```

#Anàlisi de les temperatures amb les sèries mensuals de ciclons

```

T <- T[, c(1,2,7)]
names(T) <- c("any", "mes", "temp")
par(mfrow = c(2,2))
TA <- subset(T, T$any > 1987)
TA <- subset(TA, TA$any < 2012)
TA$serie <- as.vector(ciclons.mesos.atlc)
plot.ts(TA[, c(3,4)], main = "Atlàntic")
TP <- subset(T, T$any > 2000)
TP <- subset(TP, TP$any < 2012)
TP$serie <- as.vector(ciclons.mesos.pac)
plot.ts(TP[, c(3,4)], main = "Pacífic")

```

#Imputació de la base de dades

#Separar els diferents ciclons

```

ciclo <- split(d, d$nom)
#Eliminar els ciclons amb molts NAs
nuevo <- NULL

```

```

c <- 0
for(i in 1:length(ciclo))
{
  pmv <- -100 * apply(is.na(ciclo[[i]]), 2, mean)
  if(sum(pmv < 30) == length(pmv))
  {
    c <- c + 1
    nuevo[[c]] <- ciclo[[i]]
  }
}
ciclo <- nuevo

```

```

#Imputació
library(zoo)
for(i in 1:length(ciclo))
{
  a <- as.data.frame(ciclo[[i]])
  pmv <- -100 * apply(is.na(a),2,mean)
  for(j in 1:ncol(a))
  {
    if(pmv[j] > 0)
    {
      ciclo[[i]][,j] <- na.spline(a[,j])
    }
  }
}
#Exemple de imputació
par(mfrow = c(2,1))
plot(ciclo[[2]][,11],type = "b",col = "red",main = "radi velocitat máxima del vent")
lines(ciclo[[2]][,11],type = "b")
plot(ciclo[[2]][,13],type = "b",col = "red",main = "radi isòbar més proper")
lines(ciclo[[2]][,13],type = "b")

v <- -0
for(i in 1:length(ciclo))
{
  v[i] <- -sum(as.numeric(is.na(ciclo[[i]][,13])))
}

#Tornem a recuperar la base de dades
a <- rbind(as.data.frame(ciclo[[1]]), as.data.frame(ciclo[[2]]))
for(i in 3:length(ciclo))
{
  a <- rbind(a, as.data.frame(ciclo[[i]]))
}
d <- -a
pmv <- -100 * apply(is.na(d),2,mean) #Comprovem que no hi ha NAs

#Índexs
#Arreglem les unitats de les variables
d[,7] <- -d[,7] * 0.514444 #1knot = 0.514444m/s
d[,8] <- -d[,8] * 1852 #1milla nautica = 1852 m
d[,9] <- -d[,9] * 1852 #1milla nautica = 1852 m

#Creem els índexs
pdi <- -6 * 3600 * (d[,7])^3
iei <- -6 * 3600 * ((d[,7])^3) * ((d[,8])^2)
oei <- -6 * 3600 * ((d[,7])^3) * ((d[,9])^2)
d <- data.frame(d,pdi,iei,oei)
i <- -d[,c(1,12:14)]
pdi <- aggregate(pdi ~ ID,data = i,sum)
iei <- aggregate(iei ~ ID,data = i,sum)
oei <- aggregate(oei ~ ID,data = i,sum)

```

```

i <- data.frame(pdi, iei[,2], oei[,2])
names(i) <- c("ID", "pdi", "iei", "oei")

#Diferenciem entre Atlàntic i Pacífic
atlc <- subset(d, localització == "atlàntic")
pac <- subset(d, localització == "pacífic")

#Atlàntic
pdi <- aggregate(pdi ~ ID, data = atlc, sum)
iei <- aggregate(iei ~ ID, data = atlc, sum)
oei <- aggregate(oei ~ ID, data = atlc, sum)
ia <- data.frame(pdi, iei[,2], oei[,2])
names(ia) <- c("ID", "pdi", "iei", "oei")

#Pacífic
pdi <- aggregate(pdi ~ ID, data = pac, sum)
iei <- aggregate(iei ~ ID, data = pac, sum)
oei <- aggregate(oei ~ ID, data = pac, sum)
ip <- data.frame(pdi, iei[,2], oei[,2])
names(ip) <- c("ID", "pdi", "iei", "oei")

#Estudiem la relació entre els índexs
#Atlàntic
cor(ia[, c("iei", "oei", "pdi")], use = "complete.obs")
cor(log(ia[, c("iei", "oei", "pdi")]), use = "complete.obs")
summary(lm(log(ia$iei)~log(ia$oei)))
summary(lm(log(ia$iei)~log(ia$pdi)))
summary(lm(log(ia$pdi)~log(ia$oei)))
par(mfrow = c(2,2))
plot(log(ia$iei), log(ia$oei), xlab = "log(IEI)", ylab = "log(OEI)")
plot(log(ia$iei), log(ia$pdi), xlab = "log(IEI)", ylab = "log(PDI)")
plot(log(ia$pdi), log(ia$oei), xlab = "log(PDI)", ylab = "log(OEI)")
cor(ip[, c("iei", "oei", "pdi")], use = "complete.obs")
cor(log(ip[, c("iei", "oei", "pdi")]), use = "complete.obs")

#Pacífic
summary(lm(log(ip$iei)~log(ip$oei)))
summary(lm(log(ip$iei)~log(ip$pdi)))
summary(lm(log(ip$pdi)~log(ip$oei)))
par(mfrow = c(2,2))
plot(log(ip$iei), log(ip$oei), xlab = "log(IEI)", ylab = "log(OEI)")
plot(log(ip$iei), log(ip$pdi), xlab = "log(IEI)", ylab = "log(PDI)")
plot(log(ip$pdi), log(ip$oei), xlab = "log(PDI)", ylab = "log(OEI)")

#Contruïm els gràfics de densitat
b <- -1e + 8
for(i in 1:32)
{
  b[i + 1] <- -b[i] * (10)^(1/6)
}
h <- hist(ia$pdi, breaks = b)
c <- h$counts
d <- h$density

```

```

b <- -b[-length(b)]
m <- -h$mids
g <- data.frame(d, c, b, m)
ga1 <- subset(g, g$d > 0)
plot(ga1$m, ga1$d, type = "b", log = "xy", main = "Atlántic", xlab = "PDI", ylab = "Densitat")

```

```

b <- -1e + 8
for(i in 1:25)
{
  b[i + 1] <- -b[i] * (10)^(1/6)
}
h <- hist(ip$pdi, breaks = b)
c <- h$counts
d <- h$density
b <- -b[-length(b)]
m <- -h$mids
g <- data.frame(d, c, b, m)
gp1 <- subset(g, g$d > 0)
plot(gp1$m, gp1$d, type = "b", log = "xy", main = "Pacífic", xlab = "PDI", ylab = "Densitat")

```

```

b <- -1e + 17
for(i in 1:55)
{
  b[i + 1] <- -b[i] * (10)^(1/6)
}
h <- hist(ia$iei, breaks = b)
c <- h$counts
d <- h$density
b <- -b[-(length(b) - 1)]
m <- -h$mids
g <- data.frame(d, c, b, m)
ga2 <- subset(g, g$d > 0)
plot(ga2$m, ga2$d, type = b, log = xy, main = Atlántic, xlab = IEL, ylab = Densitat) < -1e + 17

```

```

for(i in 1:49)
{
  b[i + 1] <- -b[i] * (10)^(1/6)
}
h <- hist(ip$iei, breaks = b)
c <- h$counts
d <- h$density
b <- -b[-length(b)]
m <- -h$mids
g <- data.frame(d, c, b, m)
gp2 <- subset(g, g$d > 0)
plot(gp2$m, gp2$d, type = "b", log = "xy", main = "Pacífic", xlab = "IEL", ylab = "Densitat")

```

```

b <- -1e + 19
for(i in 1:37)
{
  b[i + 1] <- -b[i] * (10)^(1/6)
}

```

```

}
h <- hist(ia$oei, breaks = b)
c <- h$counts
d <- h$density
b <- b[-1]
m <- h$mids
g <- data.frame(d, c, b, m)
ga3 <- subset(g, g$d > 0)
plot(ga3$m, ga3$d, type = "b", log = "xy", main = "Atlántic", xlab = "OEI", ylab = "Densitat")

b <- -1e + 19
for(i in 1:35)
{
  b[i + 1] <- -b[i] * (10)^(1/6)
}
h <- hist(ip$oei, breaks = b)
c <- h$counts
d <- h$density
b <- b[-1]
m <- h$mids
g <- data.frame(d, c, b, m)
gp3 <- subset(g, g$d > 0)
plot(gp3$m, gp3$d, type = "b", log = "xy", main = "Pacífic", xlab = "OEI", ylab = "Densitat")

par(mfrow = c(1,2))
plot(ga1$m, ga1$d, type = "b", log = "xy", main = "Atlántic", xlab = "PDI", ylab = "Densitat")
plot(gp1$m, gp1$d, type = "b", log = "xy", main = "Pacífic", xlab = "PDI", ylab = "Densitat")
plot(ga2$m, ga2$d, type = "b", log = "xy", main = "Atlántic", xlab = "IEI", ylab = "Densitat")
plot(gp2$m, gp2$d, type = "b", log = "xy", main = "Pacífic", xlab = "IEI", ylab = "Densitat")
plot(ga3$m, ga3$d, type = "b", log = "xy", main = "Atlántic", xlab = "OEI", ylab = "Densitat")
plot(gp3$m, gp3$d, type = "b", log = "xy", main = "Pacífic", xlab = "OEI", ylab = "Densitat")

#Càlcul del pendent
par(mfrow = c(1,2))
apdia <- lm(log(d) ~ log(m), data = ga1[-c(1:4),])
plot(log(ga1$d) ~ log(ga1$m), type = "b", main = "Atlántic", xlab = "PDI", ylab = "Densitat")
lines(log(ga1[-c(1:4),]$m), apdia$fitted.values, col = "red")
apdip <- lm(log(d) ~ log(m), data = gp1[-c(1:5),])
plot(log(gp1$d) ~ log(gp1$m), type = "b", main = "Pacífic", xlab = "PDI", ylab = "Densitat")
lines(log(gp1[-c(1:5),]$m), apdip$fitted.values, col = "red")
aieia <- lm(log(d) ~ log(m), data = ga2[-c(1:4),])
plot(log(ga2$d) ~ log(ga2$m), type = "b", main = "Atlántic", xlab = "IEI", ylab = "Densitat")
lines(log(ga2[-c(1:4),]$m), aieia$fitted.values, col = "red")
aieip <- lm(log(d) ~ log(m), data = gp2[-c(1:7),])
plot(log(gp2$d) ~ log(gp2$m), type = "b", main = "Pacífic", xlab = "IEI", ylab = "Densitat")
lines(log(gp2[-c(1:7),]$m), aieip$fitted.values, col = "red")
aoeia <- lm(log(d) ~ log(m), data = ga3[-c(1:5),])
plot(log(ga3$d) ~ log(ga3$m), type = "b", main = "Atlántic", xlab = "OEI", ylab = "Densitat")
lines(log(ga3[-c(1:5),]$m), aoeia$fitted.values, col = "red")
aoeip <- lm(log(d) ~ log(m), data = gp3[-c(1:6),])
plot(log(gp3$d) ~ log(gp3$m), type = "b", main = "Pacífic", xlab = "OEI", ylab = "Densitat")
lines(log(gp3[-c(1:6),]$m), aoeip$fitted.values, col = "red")

```

```

#Càlcul de l'interval de confiança del pendent
#Bootstrap parelles
n <- -length(ga1[-c(1:4),1])
data <- -cbind(log(ga1[-c(1:4),4]), log(ga1[-c(1:4),1]))
theta <- -summary(lm(data[,2]~data[,1]))$coefficients[2]
sdtheta <- -summary(lm(data[,2]~data[,1]))$coefficients[4]
nb <- -1000; z <- -seq(1,n); tb <- -numeric(nb)
thetab <- -numeric(nb)
for(i in 1:nb)
{
  zb <- -sample(z,n,replace = T)
  ajustb <- -lm(data[zb,2]~ data[zb,1])
  thetab[i] <- -summary(ajustb)$coefficients[2]
  sdthetab <- -summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

n <- -length(gp1[-c(1:5),4])
data <- -cbind(log(gp1[-c(1:5),4]), log(gp1[-c(1:5),1]))
theta <- -summary(lm(data[,2]~data[,1]))$coefficients[2]
sdtheta <- -summary(lm(data[,2]~data[,1]))$coefficients[4]
nb <- -1000; z <- -seq(1,n); tb <- -numeric(nb)
thetab <- -numeric(nb)
for(i in 1:nb)
{
  zb <- -sample(z,n,replace = T)
  ajustb <- -lm(data[zb,2]~ data[zb,1])
  thetab[i] <- -summary(ajustb)$coefficients[2]
  sdthetab <- -summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

n <- -length(ga2[-c(1:4),1])
data <- -cbind(log(ga2[-c(1:4),4]), log(ga2[-c(1:4),1]))
theta <- -summary(lm(data[,2]~data[,1]))$coefficients[2]
sdtheta <- -summary(lm(data[,2]~data[,1]))$coefficients[4]
nb <- -1000; z <- -seq(1,n); tb <- -numeric(nb)
thetab <- -numeric(nb)
for(i in 1:nb)
{
  zb <- -sample(z,n,replace = T)
  ajustb <- -lm(data[zb,2]~ data[zb,1])
  thetab[i] <- -summary(ajustb)$coefficients[2]
  sdthetab <- -summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

```



```

n <- length(gp2[-c(1:7),1])
data <- cbind(log(gp2[-c(1:7),4]), log(gp2[-c(1:7),1]))
theta <- summary(lm(data[,2]~data[,1]))$coefficients[2]
sdtheta <- summary(lm(data[,2]~data[,1]))$coefficients[4]
nb <- -1000; z <- seq(1,n); tb <- numeric(nb)
thetab <- numeric(nb)
for(i in 1:nb)
{
  zb <- sample(z,n,replace = T)
  ajustb <- lm(data[zb,2]~ data[zb,1])
  thetab[i] <- summary(ajustb)$coefficients[2]
  sdthetab <- summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

n <- length(ga3[-c(1:5),1])
data <- cbind(log(ga3[-c(1:5),4]), log(ga3[-c(1:5),1]))
theta <- summary(lm(data[,2]~data[,1]))$coefficients[2]
sdtheta <- summary(lm(data[,2]~data[,1]))$coefficients[4]
nb <- -1000; z <- seq(1,n); tb <- numeric(nb)
thetab <- numeric(nb)
for(i in 1:nb)
{
  zb <- sample(z,n,replace = T)
  ajustb <- lm(data[zb,2]~ data[zb,1])
  thetab[i] <- summary(ajustb)$coefficients[2]
  sdthetab <- summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

n <- length(gp3[-c(1:6),1])
data <- cbind(log(gp3[-c(1:6),4]), log(gp3[-c(1:6),1]))
theta <- summary(lm(data[,2]~data[,1]))$coefficients[2]
sdtheta <- summary(lm(data[,2]~data[,1]))$coefficients[4]
nb <- -1000; z <- seq(1,n); tb <- numeric(nb)
thetab <- numeric(nb)
for(i in 1:nb)
{
  zb <- sample(z,n,replace = T)
  ajustb <- lm(data[zb,2]~ data[zb,1])
  thetab[i] <- summary(ajustb)$coefficients[2]
  sdthetab <- summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

```

```

#Bootstrap residus
n <- -length(ga1[-c(1:4),1])
data <- -cbind(log(ga1[-c(1:4),4]), log(ga1[-c(1:4),1]))
a <- -lm(data[,2]~data[,1]);
res <- -residuals(a)
theta <- -summary(a)$coefficients[2]
sdtheta <- -summary(a)$coefficients[4]
nb <- -1000; tb <- -numeric(nb)
thetab <- -numeric(nb)
for(i in 1:nb)
{
  zb <- -sample(res,n,replace = T)
  yb <- -a$coefficient[1] + a$coefficient[2] * data[,1] + zb
  ajustb <- -lm(yb ~ data[,1])
  thetab[i] <- -summary(ajustb)$coefficients[2]
  sdthetab <- -summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

n <- -length(gp1[-c(1:5),1])
data <- -cbind(log(gp1[-c(1:5),4]), log(gp1[-c(1:5),1]))
a <- -lm(data[,2]~data[,1]);
res <- -residuals(a)
theta <- -summary(a)$coefficients[2]
sdtheta <- -summary(a)$coefficients[4]
nb <- -1000; tb <- -numeric(nb)
thetab <- -numeric(nb)
for(i in 1:nb)
{
  zb <- -sample(res,n,replace = T)
  yb <- -a$coefficient[1] + a$coefficient[2] * data[,1] + zb
  ajustb <- -lm(yb ~ data[,1])
  thetab[i] <- -summary(ajustb)$coefficients[2]
  sdthetab <- -summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

```

```

n <- -length(ga2[-c(1:4),1])
data <- -cbind(log(ga2[-c(1:4),4]), log(ga2[-c(1:4),1]))
a <- -lm(data[,2]~data[,1]);
res <- -residuals(a)
theta <- -summary(a)$coefficients[2]
sdtheta <- -summary(a)$coefficients[4]
nb <- -1000; tb <- -numeric(nb)
thetab <- -numeric(nb)
for(i in 1:nb)
{
  zb <- -sample(res,n,replace = T)

```

```

    yb <- a$coefficient[1] + a$coefficient[2] * data[,1] + zb
    ajustb <- lm(yb ~ data[,1])
    thetab[i] <- summary(ajustb)$coefficients[2]
    sdthetab <- summary(ajustb)$coefficients[4]
    tb[i] <- -(thetab[i] - theta)/sdthetab
  }
  quantile(thetab, c(0.025, 0.975))

n <- -length(gp2[-c(1: 7), 1])
data <- cbind(log(gp2[-c(1: 7), 4]), log(gp2[-c(1: 7), 1]))
a <- lm(data[, 2] ~ data[, 1]);
res <- residuals(a)
theta <- summary(a)$coefficients[2]
sdtheta <- summary(a)$coefficients[4]
nb <- -1000; tb <- numeric(nb)
thetab <- numeric(nb)
for(i in 1: nb)
{
  zb <- sample(res, n, replace = T)
  yb <- a$coefficient[1] + a$coefficient[2] * data[,1] + zb
  ajustb <- lm(yb ~ data[,1])
  thetab[i] <- summary(ajustb)$coefficients[2]
  sdthetab <- summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab, c(0.025, 0.975))

n <- -length(ga3[-c(1: 5), 1])
data <- cbind(log(ga3[-c(1: 5), 4]), log(ga3[-c(1: 5), 1]))
a <- lm(data[, 2] ~ data[, 1]);
res <- residuals(a)
theta <- summary(a)$coefficients[2]
sdtheta <- summary(a)$coefficients[4]
nb <- -1000; tb <- numeric(nb)
thetab <- numeric(nb)
for(i in 1: nb)
{
  zb <- sample(res, n, replace = T)
  yb <- a$coefficient[1] + a$coefficient[2] * data[,1] + zb
  ajustb <- lm(yb ~ data[,1])
  thetab[i] <- summary(ajustb)$coefficients[2]
  sdthetab <- summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab, c(0.025, 0.975))

n <- -length(gp3[-c(1: 6), 1])
data <- cbind(log(gp3[-c(1: 6), 4]), log(gp3[-c(1: 6), 1]))
a <- lm(data[, 2] ~ data[, 1]);
res <- residuals(a)
theta <- summary(a)$coefficients[2]
sdtheta <- summary(a)$coefficients[4]

```

```

nb <- -1000; tb <- -numeric(nb)
thetab <- -numeric(nb)
for(i in 1:nb)
{
  zb <- -sample(res,n,replace = T)
  yb <- -a$coefficient[1] + a$coefficient[2] * data[,1] + zb
  ajustb <- -lm(yb ~ data[,1])
  thetab[i] <- -summary(ajustb)$coefficients[2]
  sdthetab <- -summary(ajustb)$coefficients[4]
  tb[i] <- -(thetab[i] - theta)/sdthetab
}
quantile(thetab,c(0.025,0.975))

#Ajust: Kernel
par(mfrow = c(1,2))
iapdik <- -density(ia$pdi)
plot(ga1$d ~ ga1$m, type = "b", main = "PDI Atàntic", xlab = "PDI", ylab = "Density")
lines(iapdik$x, iapdik$y, col = "red")
ippdik <- -density(ip$pdi)
plot(gp1$d ~ gp1$m, type = "b", main = "PDI Pacífic", xlab = "PDI", ylab = "Density")
lines(ippdik$x, ippdik$y, col = "red")
iaieik <- -density(ia$iei)
plot(ga2$d ~ ga2$m, xlim = c(0,10^24), type = "b", main = "IEI Atàntic", xlab = "IEI", ylab
      = "Density")
lines(iaieik$x, iaieik$y, col = "red")
ipieik <- -density(ip$iei)
plot(gp2$d ~ gp2$m, xlim = c(0,10^22), type = "b", main = "IEI Pacífic", xlab = "IEI", ylab
      = "Density")
lines(ipieik$x, ipieik$y, col = "red")
iaoeik <- -density(ia$oei)
plot(ga3$d ~ ga3$m, xlim = c(0,10^23), type = "b", main = "OEI Atàntic", xlab = "OEI", ylab
      = "Density")
lines(iaoeik$x, iaoeik$y, col = "red")
ipoeik <- -density(ip$oei)
plot(gp3$d ~ gp3$m, xlim = c(0,10^23), type = "b", main = "OEI Pacífic", xlab = "OEI", ylab
      = "Density")
lines(ipoeik$x, ipoeik$y, col = "red")

#Test Anderson – Darling
PDIA2 <- -subset(ia, ia$pdi > 1747279327)
PDIA1 <- -subset(ia, ia$pdi > 4.641589e + 08)
PDIP2 <- -subset(ip, ip$pdi > 1.055118972e + 13)
PDIP1 <- -subset(ip, ip$pdi > 6.812921e + 08)
IEIA2 <- -subset(ia, ia$iei > 3.030056419e + 18)
IEIA1 <- -subset(ia, ia$iei > 4.641589e + 17)
IEIP2 <- -subset(ip, ip$iei > 4.862977242e + 18)
IEIP1 <- -subset(ip, ip$iei > 1.467799e + 18)
OEIA2 <- -subset(ia, ia$oei > 8.954468458e + 19)
OEIA1 <- -subset(ia, ia$oei > 1.000000e + 20)
OEIP2 <- -subset(ip, ip$oei > 2.266191063e + 20)
OEIP1 <- -subset(ip, ip$oei > 1.467799e + 20)

```

```

library(nsRFA)
gofGENPARtest(PDIA1$pd, Nsim = 2000)
gofGENPARtest(PDIA2$pd, Nsim = 2000)
gofGENPARtest(PDIP1$pd, Nsim = 2000)
gofGENPARtest(PDIP2$pd, Nsim = 2000)
gofGENPARtest(IEIA1$ei, Nsim = 2000)
gofGENPARtest(IEIA2$ei, Nsim = 2000)
gofGENPARtest(IEIP1$ei, Nsim = 2000)
gofGENPARtest(IEIP2$ei, Nsim = 2000)
gofGENPARtest(OEIA1$oei, Nsim = 2000)
gofGENPARtest(OEIA2$oei, Nsim = 2000)
gofGENPARtest(OEIP1$oei, Nsim = 2000)
gofGENPARtest(OEIP2$oei, Nsim = 2000)

#Estimació de l'exponent de potència per màxima versemblança
length(PDIA2$pd)/(sum(log(PDIA2$pd/1747279327)))
length(PDIA1$pd)/(sum(log(PDIA1$pd/4.641589e + 08)))
length(PDIP2$pd)/(sum(log(PDIP2$pd/1.055118972e + 13)))
length(PDIP1$pd)/(sum(log(PDIP1$pd/6.812921e + 08)))
length(IEIA2$ei)/(sum(log(IEIA2$ei/3.030056419e + 18)))
length(IEIA1$ei)/(sum(log(IEIA1$ei/4.641589e + 17)))
length(IEIP2$ei)/(sum(log(IEIP2$ei/4.862977242e + 18)))
length(IEIP1$ei)/(sum(log(IEIP1$ei/1.467799e + 18)))
length(OEIA2$oei)/(sum(log(OEIA2$oei/8.954468458e + 19)))
length(OEIA1$oei)/(sum(log(OEIA1$oei/1.000000e + 20)))
length(OEIP2$oei)/(sum(log(OEIP2$oei/2.266191063e + 20)))
length(OEIP1$oei)/(sum(log(OEIP1$oei/1.467799e + 20)))

```

BIBLIOGRAFIA

Disipación y Energía de los ciclones Tropicales. ajustes y test de Bondad de Ajuste, Oliver Planes. Treball de Máster 2012.

http://rammb.cira.colostate.edu/research/tropical_cyclones/tc_extended_best_track_dataset/ [24-6-1013]

<http://www.erh.noaa.gov/box/hurricane/tropicalDefinitions.shtml> [24-6-1013]

http://en.wikipedia.org/wiki/Tropical_cyclone#Tropical_depression [24-6-1013]

http://www.prh.noaa.gov/cphc/pages/FAQ/Basic_Definitions.php [24-6-1013]

<http://www.cpc.ncep.noaa.gov/data/indices/sstoi.atl.indices> [24-6-1013]

<http://banach.millersville.edu/~BobBuchanan/math375/CubicSpline/main.pdf> [24-6-1013]

<http://rss.acs.unt.edu/Rdoc/library/zoo/html/na.approx.html> [24-6-1013]

http://es.wikipedia.org/wiki/Ley_de_potencias [24-6-1013]

<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/density.html> [24-6-1013]

http://en.wikipedia.org/wiki/Anderson%E2%80%93Darling_test [24-6-1013]

<http://rss.acs.unt.edu/Rdoc/library/nsRFA/html/GOFmontecarlo.html> [24-6-1013]

Pàg. 113, Bradley Efron, Robert Tibshirani: an introduction to the bootstrap, Chapman-Hall, 1993.

Rosario Delgado, Maria Jolis, Frederic Utzet: Manderlbrot i l' atzar. Butlletí de la societat Catalana de Matemàtiques. Volum 27, n° 2, pàg 121-160, 2012.

Pàg 45, B. W. Silverman. Density Estimation for Statistics and Data Analysis, Chapman & Hall, 1992.